

## SUMMARIES

OM nucleic - nucleic search, using sw model

Title:	US-09-893-371-3
Perfect score:	3661
Sequence:	1 ggaagcgggagcgccgggtccg.....ccacctgcacctttagcca 3661

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

```
Searched:      2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280
```

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing:	Minimum Match	0%
	Maximum Match	100%
	Listing first	45 summaries

Database :

1:	gb_emb1.*
2:	gb_hlg.*
3:	gb_in.*
4:	gb_om.*
5:	gb_ov.*
6:	gb_pat.*
7:	gb_ph.*
8:	gb_pl.*
9:	gb_pr.*
10:	gb_ro.*
11:	gb_sts.*
12:	gb_sy.*
13:	gb_un.*
14:	gb_vi.*
15:	em_ba.*
16:	em_fun.*
17:	em_hum.*
18:	em_in.*
19:	em_mu.*
20:	em_om.*
21:	em_ov.*
22:	em_ov.*
23:	em_pat.*
24:	em_ph.*
25:	em_pl.*
26:	em_ro.*
27:	em_sts.*
28:	em_un.*
29:	em_vi.*
30:	em_hlg_hum.*
31:	em_hlg_iny.*
32:	em_hlg_other.*
33:	em_hlg_mus.*
34:	em_hlg_pla.*
35:	em_hlg_rod.*
36:	em_hlg_mam.*
37:	em_hlg_vrt.*
38:	em_sy.*
39:	em_hlgo_hum.*
40:	em_hlgo_mus.*
41:	em_hlgo_other.*

Pred. No. is the number of results predicted by chance to have a

Result	No.	Score	Query	Match	Length	DB	ID	Description
	1	364.9	99.7	3663	10	AY007255	AY007255 Mus musculus	
	2	3542.6	96.8	5527	10	BC021481	BC021481 Mus musculus	
	3	3382.8	88.3	3810	10	RMN96723	X74800 R.norvegicus	
	4	3097.4	84.6	3111	10	MMU967233	U967373 Mus musculus	
	5	2626.2	71.7	3384	9	HSN9701B	X98507 H.sapiens m	
	6	2595.2	70.9	3505	4	BT003420	X05420 Bos taurus	
	7	2511.2	68.6	3148	4	BT00701B	Z22852 B.taurus my	
	8	2417.6	66.0	2568	10	MMMB	X39682 M.musculus	
	9	1724.4	47.1	3288	5	RC014382	U14368 Rana catesb	
	10	1632.8	44.6	3452	5	RC0164549	U14549 Rana catesb	
	11	919.6	25.1	3445	3	AY069044	AY069044 Drosophila	
	12	916.6	25.0	3382	3	DM007956	U07596 Drosophila	
	13	749.4	20.5	116374	2	AC012834	AC012834 Drosophila	
	14	749.4	20.5	175481	3	AC010007	AC010007 Drosophila	
	15	749.4	20.5	268369	3	AC005847	AC005847 Drosophila	
	16	749.4	20.5	300412	3	AE003471	AE003471 Drosophila	
	17	585	16.0	2529	10	RMU25148	U25148 Rattus norv	
	18	584.8	16.0	2726	10	AF009960	AF009960 Mus muscu	
	19	576.6	15.7	3480	5	GD004046	U04046 Gallus dome	
	20	575.8	15.7	3779	4	BOVM74C	U02819 Bos taurus	
	21	551	15.1	3114	5	GGBBM1	X58479 Chloeken mRN	
	22	551	15.1	3126	5	AVBBM1	X51671 Avian RNA f	
	23	543.4	14.8	3132	9	AF009961	AF009961 Homo sapi	
	24	541.8	14.8	3549	9	AF127026	AF127026 Homo sapi	
	25	541.8	14.8	3590	9	AF105424	AF105424 Homo sapi	
	26	525	14.3	3607	10	RMN971A	X68199 R.norvegicus	
	27	517	14.1	3650	10	MMNHCL	X69997 M.musculus	
	28	516	14.1	4194	10	MMN9S01X	U00923 Mouse myos	
	29	507.4	13.9	3886	3	DM007595	U07595 Drosophila	
	30	501	13.7	4315	3	AY075567	AY075567 Drosophila	
	31	476.4	13.0	1900	9	AF380932	AF380932 Homo sapi	
	32	474.8	13.0	1900	9	AF380933	AF380933 Homo sapi	
	33	471.6	12.9	3762	10	RMN974	X71997 R.norvegicus	
	34	469	12.8	192364	2	AL663094	AL663094 Mus muscu	
	35	469	12.8	199345	2	OC084102	OC084102 Mus muscu	
	36	462.4	12.6	3860	9	BC028071	BC028071 Homo sapi	
	37	461	12.6	157465	10	AL591440	AL591440 Mouse DN	
	38	460.8	12.6	3617	6	AX394541	AX394541 Sequence	
	39	460.8	12.6	3617	9	HSA310570	AJ310570 Homo sapi	
	40	460.8	12.6	5347	9	AK092877	AK092877 Homo sapi	
	41	458.4	12.5	3718	5	GGBBM1A	X70400 G.gallus mR	
	42	453	12.4	3561	3	AF051353	AF051353 Acanthamo	
	43	435.6	11.9	1086	3	DR06NEMO	L13070 Drosophila	
	44	432.8	11.8	3050	9	HSN803609	ALB32202 Homo sapi	
	45	423.6	11.6	3777	10	MMN9S01NT	X57650 M.musculus	

## ALIGNMENTS

RESULT 1					
AY007255					
LOCUS	AY007255	3663 bp	mRNA	linear	ROD 01-NOV-2000
DEFINITION	Mus musculus nuclear myosin I beta mRNA, complete cds.				
ACCESSION	AY007255				
VERSION	AY007255.1	GI:11067001			
KEYWORDS	.				
SOURCE	Mus musculus.				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 3663)				
REFERENCE	Pestic-Dragovich,L., Stojiljkovic,L., Piliimonenko,A.A., Nowak,G., Ke,Y., Settlage,R.E., Shabanowitz,J., Hunt,D.F., Hozak,P. and de Laetrolle,P.				
AUTHORS					

CDS

```

Poor Match      99.7% | Score 3649; DB 10; Length 3663;
Best Local Similarity 100.0% | Pred. No. 0;
Matches 3660; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
Key 1 GCAGCGGGGCCCCCGGTCCGCACAGATGCCCTTACCGGCGCATTCGCCCTTGCGCAGTAGCGG 60
    |||||||
1 GCAGCGGGGCCCCCGGTCCGCACAGATGCCCTTACCGGCGCATTCGCCCTTGCGCAGTAGCGG 60

```

6 GGTTCAGTGAACCATGGAGAGGGCCCTTGACTGGCCAGACCGGGTAAAGGGGTACAGAACTT  
 Db 61 GGTTCAGTGAACCATGGAGAGGGCCCTTGACTGGCCAGACCGGGTAAAGGGGTACAGAACTT  
 Qy 121 TGTCTCTCTGGAGAAATTTTCAACAGTGAAGGCTCCCTCAATTTGAGAACCTCCGGGGGGCTT  
 Db 121 TGTCTCTCTGGAGAAATTTTCAACAGTGAAGGCTCCCTCAATTTGAGAACCTCCGGGGGGCTT  
 Qy 181 CCGGAGAAACCAATTTTATACCTACATCGGCTGTCTAGTCTCTGTCAATCCCTACCG  
 Db 181 CCGGAGAAACCAATTTTATACCTACATCGGCTGTCTAGTCTCTGTCAATCCCTACCG  
 Qy 241 AAGACCTTACAGTCTCAACAGCCGGCAGCATATGGAACGCTACCGTGGTGGTACGTTCTATGA  
 Db 241 AAGACCTTACAGTCTCAACAGCCGGCAGCATATGGAACGCTACCGTGGTGGTACGTTCTATGA  
 Qy 301 AGTACCAACCTCAATTTTGTTCAGATGGCTGACACTGTATACCGGGACACTGTACTATGATGA  
 Db 301 AGTACCAACCTCAATTTTGTTCAGATGGCTGACACTGTATACCGGGACACTGTACTATGATGA  
 Qy 361 TCGGGACCAAGGCAAGTGAATTTTGTGAGAGAGAGTGGGCAAGCAAGACAGAGGCCACCAA  
 Db 361 TCGGGACCAAGGCAAGTGAATTTTGTGAGAGAGAGTGGGCAAGCAAGACAGAGGCCACCAA  
 Qy 421 GAGACTCTCTCAAGTTCTATGACAGAGACCTGGCCACACCCCTGAACGGGGTGGCGGACGGC  
 Db 421 GAGACTCTCTCAAGTTCTATGACAGAGACCTGGCCACACCCCTGAACGGGGTGGCGGACGGC

Db 421 GAGATGCTCCAGTTCTTATCCAGAGACTGCCAGCCCTCGAAAGSGGGTGGCGCATGTCG 480  
 Oy 481 AGACCGCTCTGTGGAGAGCAACCCGCTGTGTAGAGCCCTTGGGAATGGCAAGACTCTCG 480  
 Db 481 AGACCGCTCTGTGGAGAGCAACCCGCTGTGTAGAGCCCTTGGGAATGGCAAGACTCTCG 540  
 Oy 541 CAACGATAACTCCAGCCGCTTGGAAAGTACATGATGTGCAAGTTTACTTCAAGGTGCG 540  
 Db 541 CAACGATAACTCCAGCCGCTTGGAAAGTACATGATGTGCAAGTTTACTTCAAGGTGCG 540  
 Oy 601 CCCCCTGGAGAGCCACATTTCTCAGTTTACTCCTGGAAAGCCCGGCTGTGCACCAAA 660  
 Db 601 CCCCCTGGAGAGCCACATTTCTCAGTTTACTCCTGGAAAGCCCGGCTGTGCACCAAA 660  
 Oy 661 TCACGGAGACGGGAATCTCCAGCTCTTAAACAGCTTACAGTGTGAGAGAGGAGGAG 660  
 Db 661 TCACGGAGACGGGAATCTCCAGCTCTTAAACAGCTTACAGTGTGAGAGAGGAGGAG 660  
 Oy 721 TCTCCCTGGCTGGGCTTGGAGAGGAACCCCAAGACTACTTGTACTGGTGAAGGGCA 720  
 Db 721 TCTCCCTGGCTGGGCTTGGAGAGGAACCCCAAGACTACTTGTACTGGTGAAGGGCA 720  
 Oy 721 TCTCCCTGGCTGGGCTTGGAGAGGAACCCCAAGACTACTTGTACTGGTGAAGGGCA 720  
 Db 781 GTGTCCAAAGCTTCTCCCATATCAAGACAGAGAGTACTGGAAAGTTATGAGAGGCGT 780  
 Oy 781 GTGTCCAAAGCTTCTCCCATATCAAGACAGAGAGTACTGGAAAGTTATGAGAGGCGT 780  
 Db 841 GTCCGTATTTGACTTCACTCAGAGATGAGAGGAGAGCTTGGTCAAGCTGCGCCACGCT 840  
 Oy 841 GTCCGTATTTGACTTCACTCAGAGATGAGAGGAGAGCTTGGTCAAGCTGCGCCACGCT 840  
 Db 901 CCTACATCTGGGGAACATCCACTTCTCTCTACAGAGAGCAATGCCAGTTACTAC 900  
 Oy 901 CCTACATCTGGGGAACATCCACTTCTCTCTACAGAGAGCAATGCCAGTTACTAC 900  
 Db 961 TGAGAACCACTCAATATCTACCAAGGCTCTTGGTGTGGAAGGTACAACTTAAAGGA 960  
 Oy 961 TGAGAACCACTCAATATCTACCAAGGCTCTTGGTGTGGAAGGTACAACTTAAAGGA 960  
 Db 1021 AGCCCTGAGCCCAAGGAAGATCATGCGCAAGGGGGGAAGAGCTCTGAGCCCACTGAACCT 1080  
 Oy 1021 AGCCCTGAGCCCAAGGAAGATCATGCGCAAGGGGGGAAGAGCTCTGAGCCCACTGAACCT 1080  
 Db 1081 TGAGAACGGCGCATATGCAAGGATGCGCTTGGCAAGGCTGTGTAACCCGAGATTAC 1140  
 Oy 1141 TGAGAACGGCGCATATGCAAGGATGCGCTTGGCAAGGCTGTGTAACCCGAGATTAC 1140  
 Db 1141 CTGGCTGTGAGAAAGATCAATAGTACTACTGGCTTAAAGAGCCGTAGAGCCCAAGCTG 1200  
 Oy 1201 GCGAAGCAACACGCTTCTGGGCTCTTGAGACTTAAAGGCTTGAAGTGTTCACACATA 1260  
 Db 1261 GCGAAGCAACACGCTTCTGGGCTCTTGAGACTTAAAGGCTTGAAGTGTTCACACATA 1260  
 Oy 1261 GCGAAGCAACACGCTTCTGGGCTCTTGAGACTTAAAGGCTTGAAGTGTTCACACATA 1260  
 Db 1321 GCTGACTCAGTGTGGAGAGAGAGATAGAGGCTTGAAGGATCGCCGAGGAACCTGT 1380  
 Oy 1381 GCTGACTCAGTGTGGAGAGAGAGATAGAGGCTTGAAGGATCGCCGAGGAACCTGT 1380  
 Db 1381 GCTGACTCAGTGTGGAGAGAGAGATAGAGGCTTGAAGGATCGCCGAGGAACCTGT 1380  
 Oy 1381 GCTGACTCAGTGTGGAGAGAGAGATAGAGGCTTGAAGGATCGCCGAGGAACCTGT 1380  
 Db 1441 CTCATCTTGATGAGAGAGTCTGAGAGTGTGAGAGAGAGGCTTGAAGGATCGAT 1440  
 Oy 1441 CTCATCTTGATGAGAGAGTCTGAGAGTGTGAGAGAGAGGCTTGAAGGATCGAT 1440  
 Db 1441 CTCATCTTGATGAGAGAGTCTGAGAGTGTGAGAGAGAGGCTTGAAGGATCGAT 1440  
 Oy 1501 GAGGTGAGAGCACTCTCAAGCCCACTTCACTTCTTGAAGCAACAGCTGCTGAGCA 1560

Db 1501 GAAGTTGGAGGACACTGTCACAGCCCACTTCTGACGACACAGCTGCTGACCA 1560  
 QY 1561 GAAGACACAGAAATCCTTAGACCGAGGGAGTTCGCTTCTGCAATTAATGAGAGGT 1620  
 Db 1561 GAAGACACAGAAATCCTTAGACCGAGGGAGTTCGCTTCTGCAATTAATGAGAGGT 1620  
 QY 1621 GACCTACAGTGTGACTGGTTCGATTAATAACAATGACCTTCCTCCGAACTGAA 1680  
 Db 1621 GACCTACAGTGTGACTGGTTCGATTAATAACAATGACCTTCCTCCGAACTGAA 1680  
 QY 1681 GGAGACCATGTGACCTCAATGAACCCCATCATGGCCCGCTCTTGGACAAAGAGGT 1740  
 Db 1681 GGAGACCATGTGACCTCAATGAACCCCATCATGGCCCGCTCTTGGACAAAGAGGT 1740  
 QY 1741 CAGTGACAAAGACCGCCAG- GACGGTGGCCACCCAGTTCAAGATGAGGCTCTCCAGCT 1799  
 Db 1741 CAGTGACAAAGACCGCCAGAGACGGTGGCCACCCAGTTCAAGATGAGGCTCTCCAGCT 1800  
 QY 1800 CGTGAGATCCTGAGGTCTAAGAGGCTGCTATATCCGGTGCATCAACCAACGAGC 1859  
 Db 1801 CGTGAGATCCTGAGGTCTAAGAGGCTGCTATATCCGGTGCATCAACCAACGAGC 1860  
 QY 1860 CAAGAGCCGGTCCCTTGTAGAGTGTCTATCCGACATCAGTGAAGTACTACTGGAGT 1919  
 Db 1861 CAAGAGCCGGTCCCTTGTAGAGTGTCTATCCGACATCAGTGAAGTACTACTGGAGT 1920  
 QY 1920 GATGAGAACTGCGCGTGCAGAGTGGCTTGTGCTATCCGCAATATGAGGCTTT 1979  
 Db 1921 GATGAGAACTGCGCGTGCAGAGTGGCTTGTGCTATCCGCAATATGAGGCTTT 1980  
 QY 1980 CCTGACAGAGTACAAAGTCACTGTGCCACAGACATGGCCCATGTGGCGAGGAGCCCA 2039  
 Db 1981 CCTGACAGAGTACAAAGTCACTGTGCCACAGACATGGCCCATGTGGCGAGGAGCCCA 2040  
 QY 2040 GGATGTTGGCGGTGGTGGTGCAGACACTGGGCTACAAAGCCAGAAAGTACAAATGGG 2099  
 Db 2041 GGATGTTGGCGGTGGTGGTGCAGACACTGGGCTACAAAGCCAGAAAGTACAAATGGG 2100  
 QY 2100 CAGGACTAAGATCTTATCCGATTCCTCCCAAGACTTATTTGGCAGAGAGTCCCTGGA 2159  
 Db 2101 CAGGACTAAGATCTTATCCGATTCCTCCCAAGACTTATTTGGCAGAGAGTCCCTGGA 2160  
 QY 2160 AGTCCGGCGGCGAGTCTAAGCCACCAAGATCCAGGGGCTTGAGGGCTTTCTATTGGCG 2219  
 Db 2161 AGTCCGGCGGCGAGTCTAAGCCACCAAGATCCAGGGGCTTGAGGGCTTTCTATTGGCG 2220  
 QY 2220 ACAGAAATTTCTCGGGTGAACGATACGCCATCTGTATCCAGTATGTTGGCTGGGAC 2279  
 Db 2221 ACAGAAATTTCTCGGGTGAACGATACGCCATCTGTATCCAGTATGTTGGCTGGGAC 2280  
 QY 2280 ACTGGGCGGAGAGGAGCCAGCAAGAGAGTGGGAGCCAGACCATCGTGCAGTCAAT 2339  
 Db 2281 ACTGGGCGGAGAGGAGCCAGCAAGAGAGTGGGAGCCAGACCATCGTGCAGTCAAT 2340  
 QY 2340 CCGTGGCTTCAATTTTGGCCATTCAACCCGCTGTGAATGCTTCTTGGACCA 2399  
 Db 2341 CCGTGGCTTCAATTTTGGCCATTCAACCCGCTGTGAATGCTTCTTGGACCA 2400  
 QY 2400 CGTGGCGGCTTATTTTGGCTTAACCTGAGGCGCAACTGCCCGGAATGTTCTGGACAC 2459  
 Db 2401 CGTGGCGGCTTATTTTGGCTTAACCTGAGGCGCAACTGCCCGGAATGTTCTGGACAC 2460  
 QY 2460 CTCTGGCCACACCCCACTGCGCTGAGAGAGGCTTCAGAAAGCTTCGGGAACTGG 2519  
 Db 2461 CTCTGGCCACACCCCACTGCGCTGAGAGAGGCTTCAGAAAGCTTCGGGAACTGG 2520  
 QY 2520 CATGAAGAAACATGTGTGAGAACTACTGCCGAGCATCAAGCCCTGAGTGAAGCAGAGCT 2579  
 Db 2521 CATGAAGAAACATGTGTGAGAACTACTGCCGAGCATCAAGCCCTGAGTGAAGCAGAGCT 2580  
 QY 2580 GCAGCAAAAAGGGGTGGCTAGTGAATTTTCAAGGGCAAGAAAGACACTACCCCGAGAG 2639  
 Db 2581 GCAGCAAAAAGGGGTGGCTAGTGAATTTTCAAGGGCAAGAAAGACACTACCCCGAGAG 2640

QY 2640 TGNCCACAGACTTTCATTAAGCACAGGCTTGGCACAGAGAGATACGCCAGAGTGTCT 2699  
 Db 2641 TGNCCACAGACTTTCATTAAGCACAGGCTTGGCACAGAGAGATACGCCAGAGTGTCT 2700  
 QY 2700 TCAATCTTTGGGCTGGAACCCATCCAGATATGCGCTGCCGTGTGTAATAATAGACCGTAA 2759  
 Db 2701 TCAATCTTTGGGCTGGAACCCATCCAGATATGCGCTGCCGTGTGTAATAATAGACCGTAA 2760  
 QY 2760 GGGTTTAAAGCCCTCGGCGGAGGCTGTGCTCAAGGCCAGTGGCTGTGTAATTTGGG 2819  
 Db 2761 GGGTTTAAAGCCCTCGGCGGAGGCTGTGCTCAAGGCCAGTGGCTGTGTAATTTGGG 2820  
 QY 2820 GGATGCTAAAGTCAACAGAGAAATGATTAATGCAACCTAACCGGAATCTGTGAGTGA 2879  
 Db 2821 GGATGCTAAAGTCAACAGAGAAATGATTAATGCAACCTAACCGGAATCTGTGAGTGA 2880  
 QY 2880 CCTGAGTGAATGACCTATTTTGTGCTTACGTGCAAGCTGAAGACACAGCAGAGGAGA 2939  
 Db 2881 CCTGAGTGAATGACCTATTTTGTGCTTACGTGCAAGCTGAAGACACAGCAGAGGAGA 2940  
 QY 2940 TGTGGTGTGAGAGTATGATGATGAGACACTAACCAAGAGGCGCTGAGTGA 2999  
 Db 2941 TGTGGTGTGAGAGTATGATGATGAGACACTAACCAAGAGGCGCTGAGTGA 3000  
 QY 3000 CCGCGTGAACATTAATCAACATCAACAGGAGCATAAAGTTTGCAGGGGTTCCAGGAG 3059  
 Db 3001 CCGCGTGAACATTAATCAACATCAACAGGAGCATAAAGTTTGCAGGGGTTCCAGGAG 3060  
 QY 3060 GGAAGGACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3119  
 Db 3061 GGAAGGACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3120  
 QY 3120 CCACGCTGTGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3179  
 Db 3121 CCACGCTGTGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3180  
 QY 3180 TCTGATCTCTGATGCTTCCCTTATGCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTT 3239  
 Db 3181 TCTGATCTCTGATGCTTCCCTTATGCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTT 3240  
 QY 3240 CTTCCAACAGGAGATCCATGAGACACCTCAAAAACCCAGCGTGCATAACTCTGCTTGTG 3299  
 Db 3241 CTTCCAACAGGAGATCCATGAGACACCTCAAAAACCCAGCGTGCATAACTCTGCTTGTG 3300  
 QY 3300 TGGCCCCCTTGGGTTGATCAGAGGAGCCAGGAGCTACCCCATGAGTGGGCGCAGGCGGG 3359  
 Db 3301 TGGCCCCCTTGGGTTGATCAGAGGAGCCAGGAGCTACCCCATGAGTGGGCGCAGGCGGG 3360  
 QY 3360 CCACACCAATTAAGAAAGCAGAGGCTGAGCAGGCGCAGGCGCAGGCGCTGCTGATGCCAA 3419  
 Db 3361 CCACACCAATTAAGAAAGCAGAGGCTGAGCAGGCGCAGGCGCAGGCGCTGCTGATGCCAA 3420  
 QY 3420 TATCTAAGCAAGGAATTTAACTGAGGTTTCTGTGAGATTTTGTGATGCTTTAAG 3479  
 Db 3421 TATCTAAGCAAGGAATTTAACTGAGGTTTCTGTGAGATTTTGTGATGCTTTAAG 3480  
 QY 3480 AAACATATTTTAAAGAAAGCATTCTTACCCTTAACACACTGATGTGTTTCCCT 3539  
 Db 3481 AAACATATTTTAAAGAAAGCATTCTTACCCTTAACACACTGATGTGTTTCCCT 3540  
 QY 3540 GCGTGAACAGGAGCAAGAAATGAACTGAAGACATGAGGCTGGGCGGGAAGGTCCTC 3599  
 Db 3541 GCGTGAACAGGAGCAAGAAATGAACTGAAGACATGAGGCTGGGCGGGAAGGTCCTC 3600  
 QY 3600 TTTCTGGCCAAACCTTCTTATTTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3659  
 Db 3601 TTTCTGGCCAAACCTTCTTATTTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3660  
 QY 3660 C 3660  
 Db 3661 C 3661

```

AGFAVRRKYEAFIQRKSLCPETWPMWAGRODQAVLVNHLGKRPPEYKKNRKTIFT
RPPATLEATEDSLEVRROSLATKIOAAMRGHNRKFLRVKRSALCISWRGRKRTIFT
KAARKMAOTIRRLTRGFLHSPRCPENAFELUDHVRAFLNLRDLPNRGLGR
PTPPALREASBELRLRLCMKMMWVKYRSNISPWMKOOLOKAVASEITKGRNVDTSW
VPLRFTSTRIETEEISPVILQSLSEPSIOTAVVVKYDRKGPKPRRLDTPSAVVI
VEDAKYKROKIDIANLJGISVSSDSILFVLHVOREDNOKGIVULOSDHIYELTFTKTA
LSADRYKMININOGSITFPAGSPGPDSDIFLHVOREDNOKGIVULOSDHIYELTFTKTA
ORIGIN 1284 a 1372 c 1444 g 1127 t

Query Match
Best Local Similarity 96.8%; Score 3542.6; DB 10; Length 5227;
Matches 3590; Conservative 0; Mismatches 24; Indels 3; Gaps 3;

QY 42 CGGCGCTGGCGAGTACCGGGGTTGAGTACCATGAGAGAGCGCTTACTGCGCGAGACC 101
DB 97 CGGCGCTGGCGAGTACCGGGGTTGAGTACCATGAGAGAGCGCTTACTGCGCGAGACC 156
QY 102 GGTAGGGGTGCAAGACTTGTCTGCTGAGAAATTTCCACAGTGAAGCTTCAATTG 161
DB 157 GGTAGGGGTGCAAGACTTGTCTGCTGAGAAATTTCCACAGTGAAGCTTCAATTG 216
QY 162 AGAAGCTCCGGGGGGGCTTCCGGGAGAACCCATTTATACCTACATCGCTCTGCTAG 221
DB 217 AGAAGCTCCGGGGGGGCTTCCGGGAGAACCCATTTATACCTACATCGCTCTGCTAG 276
QY 222 TCTGTGTCATTCCTTACCGAGACTTACAGATCCACAGCCGGCAGATATGAGAGCTTAC 281
DB 277 TCTGTGTCATTCCTTACCGAGACTTACAGATCCACAGCCGGCAGATATGAGAGCTTAC 336
QY 282 GTGTGTGCTAGTTTGTATGAGTACCACTCTTTGTCAGAGGGTGAAGCTTATAC 341
DB 337 GTGTGTGCTAGTTTGTATGAGTACCACTCTTTGTCAGAGGGTGAAGCTTATAC 396
QY 342 GGGGACTTGTACTAGAGCGTGGGAGCCAGGCAAGTATTTGTGAGAGAGTGGGGCAG 401
DB 397 GGGGACTTGTACTAGAGCGTGGGAGCCAGGCAAGTATTTGTGAGAGAGTGGGGCAG 456
QY 402 GCAGAGCAGAGCCACCAAGAGACTCTCTCAATTTATGAGAGAGCTGCGCCAGCCCTG 461
DB 457 GCAGAGCAGAGCCACCAAGAGACTCTCTCAATTTATGAGAGAGCTGCGCCAGCCCTG 516
QY 462 AACGGGGTGGGAGTGGGAGAGCCGCTGTTGAGAGCAACCCCGTTAAGAGCCCTTG 521
DB 517 AACGGGGTGGGAGTGGGAGAGCCGCTGTTGAGAGCAACCCCGTTAAGAGCCCTTG 576
QY 522 GAATGCGAAGACTCTCGCAAGCATTAATCTCAAGCCGTTTGAAGATGATGATGTC 581
DB 577 GAATGCGAAGACTCTCGCAAGCATTAATCTCAAGCCGTTTGAAGATGATGATGTC 636
QY 582 AGTTTGACTTCAAGGGTGGCCCGTGGAGAGCCATTTCACTTCTCTGGAAGT 641
DB 637 AGTTTGACTTCAAGGGTGGCCCGTGGAGAGCCATTTCACTTCTCTGGAAGT 696
QY 642 CCGGGTGTGTGCAACAAATTCAGGAGAGCGGAACCTTCCACTTTTACCAAGTACTG 701
DB 697 CCGGGTGTGTGCAACAAATTCAGGAGAGCGGAACCTTCCACTTTTACCAAGTACTG 756
QY 702 AGGGGGGAGAGAGAGACTCTCGCTGGCTGGGCTTGAAGAGGAGAACCCCAAGCTACT 761
DB 757 AGGGGGGAGAGAGAGACTCTCGCTGGCTGGGCTTGAAGAGGAGAACCCCAAGCTACT 816
QY 762 TGTACCTGTGTAAGGGCAATGTGCAAGAGCTCTCTCATTAACGAGAGTACTG 821
DB 817 TGTACCTGTGTAAGGGCAATGTGCAAGAGCTCTCTCATTAACGAGAGTACTG 876
QY 822 AGTTTATGAGAGAGGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 881
DB 877 AGTTTATGAGAGAGGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 936
QY 882 TCAGATCTGTGCGCAGGCTGTCAATCTGAGGCAACATCTTGTGCTGAGAGGAGACA 941
DB 937 TCAGATCTGTGCGCAGGCTGTCAATCTGAGGCAACATCTTGTGCTGAGAGGAGACA 996

```

RESULT 2  
BC021481  
LOCUS  
DEFINITION Mus musculus, myosin Ic, clone MGC:36155 IMAGE:5344331, mRNA,  
complete cds.  
ACCESSION BC021481  
VERSION BC021481.1 GI:18204813  
KEYWORDS MGC.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Strausberg, R.  
Direct Submission  
Submitted (14-JAN-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [gcapps-remail.nih.gov](mailto:gcapps-remail.nih.gov)  
Tissue Procurement: Jeffrey Green M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL),  
Sequencing Center (MGC), National Institutes of Health Intramural  
Genetics Laboratory, Maryland.  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc-mgc@nsl.nih.gov](mailto:nisc-mgc@nsl.nih.gov)  
Akhter, N., Ayle, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
Blakesley, R.W., Boufford, G.G., Green, K., Brinkley, C., Brooks, S.,  
Dierlich, N.L., Granito, S., Guan, X., Gupta, J., Haghghigh, P.,  
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Lario, P., Legaspi, R.,  
McDowell, J., Pearson, R., Maskert, B., Masruran, S.D., McCloskey, J.C.,  
Tsurgoun, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
Young, A., Zhang, L.-H. and Green, E.D.

FEATURES  
source  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 55 Row: C Column: 9  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 6678985.  
Location/Qualifiers  
1. 5227  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/map="FVB/N"  
/clone="MGC:36155 IMAGE:5344331"  
/tissue="Mammary tumor, C3(1)-Tag model. Infiltrating  
ductal carcinoma. 5 month old virgin mouse."  
/clone\_id="NCI\_CGAP\_Mam6"  
/lab\_host="DH10B"  
/note="Vector: PCMV-SPORT6"  
129. 3215  
/codon\_start=1  
/product="myosin Ic"  
/protein\_id="AAH21481.1"  
/db\_xref="gi:18204814.1"  
/db\_xref="locusid:17913"  
/translation="MESATITANDRGVQVDVPLENFTSEAFTEMLRRFRRENLYTY  
ISGSGAGCTEATKRLQYATCTCPAPRGAVVDRRLFAVADVYALTRERDAVY  
RLGKRNVDVDFKGVGGHILSYLLEKSHVYONHNGENHVEFQLEGGEBETIS  
NLGNIHFPADEDSNAOVCTAVSSINDSPMKVWRALSVIDFTDEVEDLSTVAS  
NLEQAVYARDALAKAVYSRTPTMLVRKINSLSAKDAESPSTRTTIVIGLELSEL  
FOHNSFEQCTICNEKLOQFTELILKSQDEYBAEGJAMEPQYVNNKTIQDVEE  
KFGIISTIDECRLRGVATDLELTKSDIVYKPHPLTHKLEADPTKRSIDRGTFR  
LHVAGPTVTSVTEGLRNNDLFRPLKTKTSSNNPIMACCPEKSLSDKRPETVA  
TQFKMSLIQLEILRSKPEYIIRCIPNDAKQPGRFDEVILRHQVITGLMELNLR

CDS  
129. 3215  
/product="myosin Ic"  
/protein\_id="AAH21481.1"  
/db\_xref="gi:18204814.1"  
/db\_xref="locusid:17913"  
/translation="MESATITANDRGVQVDVPLENFTSEAFTEMLRRFRRENLYTY  
ISGSGAGCTEATKRLQYATCTCPAPRGAVVDRRLFAVADVYALTRERDAVY  
RLGKRNVDVDFKGVGGHILSYLLEKSHVYONHNGENHVEFQLEGGEBETIS  
NLGNIHFPADEDSNAOVCTAVSSINDSPMKVWRALSVIDFTDEVEDLSTVAS  
NLEQAVYARDALAKAVYSRTPTMLVRKINSLSAKDAESPSTRTTIVIGLELSEL  
FOHNSFEQCTICNEKLOQFTELILKSQDEYBAEGJAMEPQYVNNKTIQDVEE  
KFGIISTIDECRLRGVATDLELTKSDIVYKPHPLTHKLEADPTKRSIDRGTFR  
LHVAGPTVTSVTEGLRNNDLFRPLKTKTSSNNPIMACCPEKSLSDKRPETVA  
TQFKMSLIQLEILRSKPEYIIRCIPNDAKQPGRFDEVILRHQVITGLMELNLR



QY 942 GCAATGCCAGTTACTACTGAGAACACAGCTCAAAATATCTGACACAGGCTCCTTGGTGG 1001  
 Db 997 GCAATGCCAGTTACTACTGAGAACACAGCTCAAAATATCTGACACAGGCTCCTTGGTGG 1056  
 QY 1002 AAGGTACAACTTATAGGAGAGCCCTGACCCACAGAGATCATCTGCCCAAGGGGAGAGC 1061  
 Db 1057 AAGGTACAACTTATAGGAGAGCCCTGACCCACAGAGATCATCTGCCCAAGGGGAGAGC 1116  
 QY 1062 TCCTGAGCCCACTGAACTTGAACAGGGGGCATATGCAAGGATGCGGCTTGGCAAGGGCTG 1121  
 Db 1117 TCCTGAGCCCACTGAACTTGAACAGGGGGCATATGCAAGGATGCGGCTTGGCAAGGGCTG 1176  
 QY 1122 TGTACAGCCGAGCATTTCACTGCTGCTGCTGAGAAAGATCAATAGTACTGAGCTTAAAG 1181  
 Db 1177 TGTACAGCCGAGCATTTCACTGCTGCTGCTGAGAAAGATCAATAGTACTGAGCTTAAAG 1236  
 QY 1182 ACGCTGAGAGCCCACTGAGCTGGGAGACACACCGTTCTTGGGCTCCTGGACATTTACGGCT 1241  
 Db 1237 ACGCTGAGAGCCCACTGAGCTGGGAGACACACCGTTCTTGGGCTCCTGGACATTTACGGCT 1296  
 QY 1242 TTGAAGTGTTCAGCATTAACACACTTGAAGCATTTGATGCACTACTGACATGAGAGC 1301  
 Db 1297 TTGAAGTGTTCAGCATTAACACACTTGAAGCATTTGATGCACTACTGACATGAGAGC 1356  
 QY 1302 TGCAGCAGCTTCTTCATCGAGCTGACTCTCAAGTCGAGAGAGAGAAATACAGAGCTGAG 1361  
 Db 1357 TGCAGCAGCTTCTTCATCGAGCTGACTCTCAAGTCGAGAGAGAGAAATACAGAGCTGAG 1416  
 QY 1362 GCATGCGCTGGAGACCTGCTCCAGTACTTCAACACACATCATCTGTGACCTGTAGAGC 1421  
 Db 1417 GCATGCGCTGGAGACCTGCTCCAGTACTTCAACACACATCATCTGTGACCTGTAGAGC 1476  
 QY 1422 AGAAGTTCAAGGGGATCATCTCTCATCTTGGATGAAGAGAGGCTGCTGGGAGAGGCA 1481  
 Db 1477 AGAAGTTCAAGGGGATCATCTCTCATCTTGGATGAAGAGAGGCTGCTGGGAGAGGCA 1536  
 QY 1482 CGGACCTGACCTTTCTTGGAGAAAGTTGAGAGACATGTCAGGCCCACTCTACTTCTGA 1541  
 Db 1537 CGGACCTGACCTTTCTTGGAGAAAGTTGAGAGACATGTCAGGCCCACTCTACTTCTGA 1596  
 QY 1542 CGGACACGCTGCTGACCAAGAGACCAAGAAATCCCTAGACCGAGGGAGTTCCGCTTC 1601  
 Db 1597 CGGACACGCTGCTGACCAAGAGACCAAGAAATCCCTAGACCGAGGGAGTTCCGCTTC 1656  
 QY 1602 TGCATTATGCTGAGAGAGTACCTACGTGACGTGAGGTTCTTGTGATTAATAACATGAGC 1661  
 Db 1657 TGCATTATGCTGAGAGAGTACCTACGTGACGTGAGGTTCTTGTGATTAATAACATGAGC 1716  
 QY 1662 TGCCTTCCGGAACCTGAAAGAGACCATGTCAGCTCAATGAACCCCATCATGAGCCAGT 1721  
 Db 1717 TGCCTTCCGGAACCTGAAAGAGACCATGTCAGCTCAATGAACCCCATCATGAGCCAGT 1776  
 QY 1722 GCTTTCACAGAGTACGTCACTGACACAGAGAGGCGCAG-GACGCTGGCCACCCAGTTCA 1780  
 Db 1777 GCTTTCACAGAGTACGTCACTGACACAGAGAGGCGCAGAGAGGCTGGCCACCCAGTTCA 1836  
 QY 1781 AGATGAGCTCCTGAGGCTGAGAGATCCTGAGGCTTAAGGAGCTGCTATATCCGT 1840  
 Db 1837 AGATGAGCTCCTGAGGCTGAGAGATCCTGAGGCTTAAGGAGCTGCTATATCCGT 1896  
 QY 1841 GCATCAAGCCAAACGAGCCGCAAGAGAGCGGCTGCTTTCATGAGTGTGCTATCCGACATC 1900  
 Db 1897 GCATCAAGCCAAACGAGCCGCAAGAGAGCGGCTGCTTTCATGAGTGTGCTATCCGACATC 1956  
 QY 1901 AGGTAAAGTACCTGGAGTATGAGAGATCTGCGCTGGCGCAGAGCTGGCTTGGCTATC 1960  
 Db 1957 AGGTAAAGTACCTGGAGTATGAGAGATCTGCGCTGGCGCAGAGCTGGCTTGGCTATC 2016  
 QY 1961 GTTCGAAATATGAGCTTCTGAGAGAGTACAGTCACTGAGCCCAAGAGATGGCCCA 2020  
 Db 2017 GTTCGAAATATGAGCTTCTGAGAGAGTACAGTCACTGAGCCCAAGAGATGGCCCA 2076

QY 2021 TGTGGGACAGAGCGGCCCCAGGATGCTGTGGCCGTGTTGTTGTCAGACACCTCGGCTACAGC 2080  
 Db 2077 TGTGGGACAGAGCGGCCCCAGGATGCTGTGGCCGTGTTGTTGTCAGACACCTCGGCTACAGC 2136  
 QY 2081 CAGAGAGTACAAATATGGGACAGACTTAAGATCTTCAATCCGATTTCCCAAGACTTATTGG 2140  
 Db 2137 CAGAGAGTACAAATATGGGACAGACTTAAGATCTTCAATCCGATTTCCCAAGACTTATTGG 2196  
 QY 2141 CCACAGAGAGCTCCCTGGAAGTCCGGCGCAGAGTACCTCCACCAAGATTCAGAGCGGCT 2200  
 Db 2197 CCACAGAGAGCTCCCTGGAAGTCCGGCGCAGAGTACCTCCACCAAGATTCAGAGCGGCT 2256  
 QY 2201 GGAGGGGCTTTCATTTGGGACAGAAATTTCTCCGGGTGAAGGATCAGCATCTGTATCC 2260  
 Db 2257 GGAGGGGCTTTCATTTGGGACAGAAATTTCTCCGGGTGAAGGATCAGCATCTGTATCC 2316  
 QY 2261 AGTCATGTTGGGCTGGGACACTGGGGCCGAGAGAGGACACCAAGAGAAAGTGGGACGCC 2320  
 Db 2317 AGTCATGTTGGGCTGGGACACTGGGGCCGAGAGAGGACACCAAGAGAAAGTGGGACGCC 2376  
 QY 2321 AGACCATCCGTGACATCCCGTGGCTTATTTTGGGCAATTTGCGCATTCACCCGGGCTGAGA 2380  
 Db 2377 AGACCATCCGTGACATCCCGTGGCTTATTTTGGGCAATTTGCGCATTCACCCGGGCTGAGA 2436  
 QY 2381 ATGCTTCTTCTTGGACACACGTGGCGCCTCATTTTGTCTTAACCTGAGCGGCAACTGC 2440  
 Db 2437 ATGCTTCTTCTTGGACACACGTGGCGCCTCATTTTGTCTTAACCTGAGCGGCAACTGC 2496  
 QY 2441 CCGGGAATTTTGGACACACTCTGGCCCAACACCCCACTGCTGCTGAGAGAGGCTCAG 2500  
 Db 2497 CCGGGAATTTTGGACACACTCTGGCCCAACACCCCACTGCTGCTGAGAGAGGCTCAG 2556  
 QY 2501 AACTGCTACGGGACCTGCTCATGAAAGATGATGTAAGTACTGCTGGAGATCAGCC 2560  
 Db 2557 AACTGCTACGGGACCTGCTCATGAAAGATGATGTAAGTACTGCTGGAGATCAGCC 2616  
 QY 2561 CTGAGTGAAGACGACGCTGACAGCAAAAGCGGTGCTAGTGAATTTTCAAGGGCAAGA 2620  
 Db 2617 CTGAGTGAAGACGACGCTGACAGCAAAAGCGGTGCTAGTGAATTTTCAAGGGCAAGA 2676  
 QY 2621 AAGACACTTACCCCAAGAGTGTCCCAAGACTTCTTATTTAGCACAGGCTTGGACAGAGG 2680  
 Db 2677 AAGACACTTACCCCAAGAGTGTCCCAAGACTTCTTATTTAGCACAGGCTTGGACAGAGG 2736  
 QY 2681 AGATCAAGCCCAAGAGTGTCTCAATCCTTGGGCTGTAACCACTGATGCTGCTCCGG 2740  
 Db 2737 AGATCAAGCCCAAGAGTGTCTCAATCCTTGGGCTGTAACCACTGATGCTGCTCCGG 2796  
 QY 2741 TGTAAATATACGACGTAAGGGTTACAAAGCTCGCCCGGACGCTGCTACGCCCA 2800  
 Db 2797 TGTAAATATACGACGTAAGGGTTACAAAGCTCGCCCGGACGCTGCTACGCCCA 2856  
 QY 2801 GTGCTGTGTCATTTGTGAGAGTCTTAAAGTCAAGCAGAGAAATGATTTGCCAACCTAA 2860  
 Db 2857 GTGCTGTGTCATTTGTGAGAGTCTTAAAGTCAAGCAGAGAAATGATTTGCCAACCTAA 2916  
 QY 2861 CCGGAATCTCTGACGTACCTGAGAGTACCTTATTTGCTTCAAGGAGAGGAG 2920  
 Db 2917 CCGGAATCTCTGACGTACCTGAGAGTACCTTATTTGCTTCAAGGAGAGGAG 2976  
 QY 2921 ACAACAAGCAGAGAGAGTGTGCTGCTCAGAGTATGATGATCGAGACACTAACCA 2980  
 Db 2977 ACAACAAGCAGAGAGAGTGTGCTGCTCAGAGTATGATGATCGAGACACTAACCA 3036  
 QY 2981 AGAGGCGCTCAGTGTGACCGGCTGACCAATATCAACATCAACAGGCGAGCATTAAGT 3040  
 Db 3037 AGAGGCGCTCAGTGTGACCGGCTGACCAATATCAACATCAACAGGCGAGCATTAAGT 3096  
 QY 3041 TTGCAAGGGGCTCAGGACAGGAGCGCATTTGACTTCAATCGGGCTCAGAGCTTCTCA 3100  
 Db 3097 TTGCAAGGGGCTCAGGACAGGAGCGCATTTGACTTCAATCGGGCTCAGAGCTTCTCA 3156  
 QY 3101 TCACCAAGGCTTAAGATGCGCACCTGGCTGTGGTGGCCCAAGGCTGAATTTCTGGTGAT 3160

```

Db 3157 TCACCAAGCTAAGATGCCACCTGCTGTGTGCGCCCAAGGCTCAATTTCTGCTGAT 3216
Oy 3161 GAAGCTCGCGTGAGCGGCTCTGACTGCTGATGCTTCCCTAGTCCCGCCCTCC 3216
Db 3217 GAAGCTCGCGTGAGCGGCTCTGACTGCTGATGCTTCCCTAGTCCCGCCCTCC 3220
Oy 3221 GACTTACCAAACTCAAGCTTCCAAAGAGGATTCATGAGACACCTCAAAACCGCT 3220
Db 3277 CAGTTACCAAACTCAAGCTTCCAAAGAGGATTCATGAGACACCTCAAAACCGCT 3276
Oy 3281 GCAAACTCGCTGCTTCTGCTGCGCCCTCTGAGTGTATGAGAGCCCTCAAAACCGCT 3280
Db 3336 GCAAACTCGCTGCTTCTGCTGCGCCCTCTGAGTGTATGAGAGCCCTCAAAACCGCT 3335
Oy 3341 ATGAGTGGCCGAGCGCGGCGCACACCAATAGAAAAGCAGAGCGCTGAGCGCCAGCC 3340
Db 3396 ATGAGTGGCCGAGCGCGGCGCACACCAATAGAAAAGCAGAGCGCTGAGCGCCAGCC 3395
Oy 3401 GCCCTCTGCTGATGCCCAATATCTAAGACAAGGATTTTAACTGAGGTTTCTGAGCA 3455
Db 3456 GCCCTCTGCTGATGCCCAATATCTAAGACAAGGATTTTAACTGAGGTTTCTGAGCA 3460
Oy 3461 TTTTGTGATGCTTATAGAAACTATTTTAAAGAAAGCCATTTTCTACCTAAACAC 3515
Db 3516 TTTTGTGATGCTTATAGAAACTATTTTAAAGAAAGCCATTTTCTACCTAAACAC 3520
Oy 3521 ACTGATGCTTATAGAAACTATTTTAAAGAAAGCCATTTTCTACCTAAACAC 3575
Db 3576 ACTGATGCTTATAGAAACTATTTTAAAGAAAGCCATTTTCTACCTAAACAC 3580
Oy 3581 GCTGGCTGGAAGGCTCTCTCTTCTGCAACAGGCGCAAGGATTTTAACTGAGGTTTCTGAGCA 3635
Db 3636 GCTGGCTGGAAGGCTCTCTCTTCTGCAACAGGCGCAAGGATTTTAACTGAGGTTTCTGAGCA 3639
Oy 3640 ATCCACCTGCACCTTTT 3656
Db 3696 ATCCACCTGCACCTTTT 3712

```

## RESULT 3

```

LOCUS RNMNR2 3810 bp mRNA linear ROD 22-SEP-1993
DEFINITION R. norvegicus MYR2 mRNA for myosin I heavy chain.
ACCESSION X74800.1 GI:400428
VERSION myosin I heavy chain, MYR2 gene.
KEYWORDS Rattus norvegicus.
SOURCE Rattus norvegicus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus;
Rattus:
1 (bases 1 to 3810)
AUTHORS Ruppert, C.
TITLE Direct Submision
JOURNAL Submitted (26-AUG-1993) C. Ruppert, Friedrich-Miescher-Laboratorium
der, Max-Planck-Gesellschaft, 72076 Tuebingen, Spemannstr. 37-39,
2 (bases 1 to 3810)
REFERENCE Ruppert, C., Godel, J., Reinhard, J. and Baehler, M.
AUTHORS MYR-2, a novel class-I Myosin Identified in rat brain
TITLE Unpublished
JOURNAL Location/Qualifiers
FEATURES
source
1. 3810
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone_lib="M115"-2,35-1,X33B"
/dev_stage="adult"
/label="3353
/gene="MYR 2"
/label="3353
/gene="MYR 2"

```

```

/codon_start=1
/product="myosin I heavy chain"
/protein_id="CAA52807.1"
/db_xref="GI:400429"
/db_xref="SPRMBL:063355"
/translation="MESALTRADRVYQDFVLENTSPAENLRRREMLITTY
IGPVLAVNPVRLQIYTRHMPRYKVSFVEVPHLPAVDVYTRRLTRERDQVM
ISGESGATETATKRLQFATGCPAPBEGGAVRRLQSPVTRVRLTRERDQVM
SRGRTRMVOEDFGAPVGGHITLSTLEKESRVHONHEERHLEAFNAKTLRNS
RGLERNPQSYLVKQCCAKVSSINDSKVMKFLVRLHRIIAKGEELLISPL
LHAGEVTVSYTGELDKNDLFRNLKEMSCSTPIMAOCFDSKSLDGRER
TOFMSILQIVELTSLKSDPEWPTWGRDQVAVVLRHLYGKPREYMGRTKLEI
AGPARRRYEARLQNTKSLCEPWPWPTWGRDQVAVVLRHLYGKPREYMGRTKLEI
RPFKTLFATEDSLERKSLATKIOAMGCPAPBEGGAVRRLQSPVTRVRLTRERDQVM
KAARKMAOITRLRLINGFLIRHAPRCPEFFLDHVTSLFLRLROLRRLVLTSM
VRLPFLSTRLTEISLPRVLOALGSPDIOAPVVKYDRKGYDPRROLLTPSAVI
VEDAKVKKORIDVNLGTISVSLSDSLFVLRVEDSKDRGDVLDSDHVTETLKTA
LSADRVNININOGSITPFGGRCRGLITDFSGSELLITRAKNGHLAVARLNSR"
BASE COUNT 905 a 1040 c 1084 g 781 t
ORIGIN

```

```

Query Match
Best Local Similarity 88.3%; Score 3232.8; DB 10; Length 3810;
Matches 3398; Conservative 0; Mismatches 182; Indels 6; Gaps 5;

```

```

Oy 1 GGAGCGGCGCGCGCGGCTCGGAGATGCGCTACCGGGCATGCGCGCGCTGCGAGTGACGG 60
Db 196 GGAGCGGCGCGCGCGGCTCGGAGATGCGCTACCGGGCATGCGCGCGCTGCGAGTGACGG 60
Oy 61 GCTTCAAGTACATGAGAGCGCGCTTACTGCGCGAGAGCGGGTGGCGAGACTT 120
Db 254 GGTACGAGTACATGAGAGCGCGCTTACTGCGCGAGAGCGGGTGGCGAGACTT 120
Oy 121 TGTCTGCTGGAGAAATTTACAGAGTGGCTGCTTACTGAGAGCTGCGAGACTT 313
Db 314 TGTCTGCTGGAGAAATTTACAGAGTGGCTGCTTACTGAGAGCTGCGAGACTT 313
Oy 181 CCGGAGAGACCTATTTATACCTATGAGAGCTGCTTACTGAGAGCTGCGAGACTT 373
Db 374 CCGGAGAGACCTATTTATACCTATGAGAGCTGCTTACTGAGAGCTGCGAGACTT 373
Oy 241 AGACCTACAGATCTACAGCGCGAGCATATGAGAGCTACCGGTGCTCATGATCCCTACCG 240
Db 434 GGACCTACAGATCTACAGCGCGAGCATATGAGAGCTACCGGTGCTCATGATCCCTACCG 433
Oy 301 AGTACACCTGATTTTGTGAGAGTGGCTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 493
Db 494 AGTACACCTGATTTTGTGAGAGTGGCTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 493
Oy 361 TCGGAGACAGAGCTGATGATTTTGTGAGAGTGGCGAGGCAAGACAGAGGCGACACAA 420
Db 554 TCGGAGACAGAGCTGATGATTTTGTGAGAGTGGCGAGGCAAGACAGAGGCGACACAA 420
Oy 421 GAGAGTGTCTCAGATTCTATGAGAGCTGCGACCGCTGAGAGGCGGCGAGCGCGAGTGG 613
Db 614 GAGAGTGTCTCAGATTCTATGAGAGCTGCGACCGCTGAGAGGCGGCGAGCGCGAGTGG 613
Oy 481 AGACCGCTGTTGAGAGACCGCGCTGTTAGAGCGCTTGGAGATATGCCAAGACTCTCG 673
Db 674 AGACCGCTGTTGAGAGACCGCGCTGTTAGAGCGCTTGGAGATATGCCAAGACTCTCG 673
Oy 541 CAAGATTAATCTCAGCGCTTGGAAAGTACATGATGCGAGTTGACTTCAAGGGTGC 600
Db 734 CAAGATTAATCTCAGCGCTTGGAAAGTACATGATGCGAGTTGACTTCAAGGGTGC 600
Oy 601 CCGGTTGGAGCGCAATTCCTGAGTAAAGTCCCGGGTGGTGGACCAAA 660
Db 794 CCGTGGAGCGCAATTCCTGAGTAAAGTCCCGGGTGGTGGACCAAA 660

```

QY	1741	CAGGACAAAGAGCGGCGAG - GACGGGGCCACCCGATCAAGATGAGGCTCCGACCT	1799
Db	1934	CAGTACAAAGAGCGGCGAAGACGGGGCCACCAGTTCAAGATGAGGCTCTTCCACCT	1993
QY	1800	CGTGGAGTCTCTGAGGCTCTAAGAGGCTGGCTATATCCGGTGCATCAAGCCAAACGAGCG	1859
Db	1994	AGTGAAGATCTGTAGGCTCTAAGGAGCTGTTATATCCGGTGCATCAAGCCAAATGAGCG	2053
QY	1860	CAAGCAGCGGGGCTGCTTGTATGATGAGTCTCATCCGACATCAGGTGAAGTACCTGGGACT	1919
Db	2054	CAAACAGCGGGGCTGCTTGTATGATGAGGCTCATCCGACACCAAGTGAAGTACCTGGGACT	2113
QY	1920	GATGAGAAATCTGGCGCTGGGCGAGAGCTGGCTTGGCTATGCTGCCAAATATAGAGCTTT	1979
Db	2114	TATGAGAAATCTGGGTGTGGCGAGAGCTGGCTTGGCTATGCTGCCAAATATAGAGCTTT	2173
QY	1980	CCCTCAGAGGATCAAGTCACTGTGCCAGAGCATGGCCATGTGGCGAGCGGCCCA	2039
Db	2174	CCTCAGAGGATCAAGTCACTGTGCCAGAGCATGGCCATGTGGCGAGCGGCCCA	2233
QY	2040	GGATGTGTGGCGCTGTTGTTGTCAGACACTTGGCTACAAGCCAGAAGAGTACAAATAGG	2099
Db	2234	GGATGTGTGGCGCTGTTGTTGTTGTCAGACACTTGGCTACAAGCCAGAAGAGTACAAATAGG	2293
QY	2100	CAGAGCTAAGTCTTCAATCCGATTTCCCCAAAGACCTTATTTGGCACAGAGAGTCCCTGGA	2155
Db	2294	CAGAGCTAAGTCTTCAATCCGATTTCCCCAAAGACCTTGTTTGGCACAGAGAGTCCCTGGA	2353
QY	2160	AGTCCGGCGGACAGTCTTACCCACCAAGATCCAGCGGCGCTGAGAGGCGCTTCATTGGCG	2219
Db	2354	AGTCCGAGCGGACAGTCTTACCCACCAAGATCCAGCGGCGCTGAGAGGCGCTTCATTGGCG	2413
QY	2220	ACAGAAATTTTCTCCGGGTGAAGCGATCAGCATCTGTATCCAGTCAATGGTGGCGTGGCAC	2279
Db	2414	GCAGAAATTTTCTCCGGGTGAAGCGATCAGCATCTGTATCCAGTCAATGGTGGCGTGGCAC	2473
QY	2280	ACTGGGCGGAGGAAAGCAGCCCAAGAGAGTGGGACGCCAGACCATCTCCGTCACATAT	2339
Db	2474	ACTGGGCGGAGGAAAGCAGCCCAAGAGAGTGGGACGCCAGACCATCTCCGTCACATAT	2533
QY	2340	CCGTGGCTTCAATTTTGGGCAATTCACCCCGGTGCCCTGAGAAATGCCCTTCTTGGACCA	2399
Db	2534	CCGTGGCTTCAATTTTGGGCTATGACACCCCGGTGCCCTGAGAAATGCCCTTCTTGGACCA	2593
QY	2400	CGTGGCGGCTCAATTTTGTCTTAACTGAGGCGGCAACTGCCCCGGAATGTTCTGGACAC	2459
Db	2594	CGTGGCGGCTCAATTTTGTCTTAACTGAGGCGGCAACTGCCCCGGAATGTTCTGGACAC	2653
QY	2460	CTCTGGGCGCCACACCCCGACCTGCGCTGAGAGAGGCGCTCAGAACTGCTACGGGAACGTG	2519
Db	2654	CTCTGGGCGCCACACCCCGACCTGCGCTGAGAGAGGCGCTCAGAACTGCTACGGGAACGTG	2713
QY	2520	CATGAAGAACAATGGTGTGGAAGTACTGCCGAGCATCAGCCCTGAGTGAAGACACAGCT	2579
Db	2714	CATGAAGAACAATGGTGTGGAAGTACTGCCGAGATATCAGCCCTGAGTGAAGACACAGCT	2773
QY	2380	GCAGCAAAAGCGGTGACTAGTGAATTTTCAAGGGCAAGAGAGACACTACCCCCAGAG	2639
Db	2774	GCAGCAAAAGCGGTGACTAGTGAATTTTCAAGGGCAAGAGAGACAAATTACCCCCAGAG	2833
QY	2640	TGTCTCCCAAGACTCTTCAATTAGCACAGCGGCTGGCACAGAGAGATACGCCCCAGAGTGT	2699
Db	2834	TGTCTCCCAAGACTCTTCAATCAGCACACAGCGCTGGCACAGAGAGATACGCCCCAGAGTGT	2893
QY	2700	TCAATCTTGGGCTCTGTAACCCATCCAGTGTGCGGAGGCCCGGTGTAATAATGACACGTAA	2759
Db	2894	TCAAGGCTTGGGCTCTGTAACCCATCCAGTGTGAGTGTCTTGTGTAAATATGACACGTAA	2953
QY	2760	GGGTACCAAGCCTGGCCCCCGGAGGCTGTCTCACGCCAGAGTGTGGTTCATTGTGGA	2819
Db	2954	GGGTACCAAGCCTGTTCGCCGGAGGTGTCTCACACCCAGAGTGTGGTTCATTGTGGA	3013
QY	2820	GGATGCTAAAGTCAAGCAGAGAAATGTGATTTATGCCAACTAACCGGAATCTCTGTAGTAG	2879



exon	999..1118 /gene="M1D" /number=9
exon	1119..1201 /gene="M1D" /number=10
exon	1202..1307 /gene="M1D" /number=11
exon	1308..1388 /gene="M1D" /number=12
exon	1389..1480 /gene="M1D" /number=13
misc_feature	1481..3110 /gene="M1D" /note="genomic structure not determined"
BASE COUNT	744 a 839 c 894 g 634 t
ORIGIN	
Query Match	84.6%; Score 3097.4; DB 10; Length 3111; Best Local Similarity 99.9%; Pred. No. 0;
Matches 3109; Conservative	0; Mismatches 1; Indels 1; Gaps 1;
OY	63 TTCCGATACCATGGAGACGCCCTTGACTGCCCGAGACC GGSTAGGGGTGCAGACTTGG 122
Db	1 TACGAGTAGCACCTGGAGAGCGCCTTGACTGCCCGAGACCGGGGTAGGGGTGCAGACTTTGG 60
OY	123 TCCTTCGTGGAGAAATTTCACACAGTAGAGGCTGCCTTCATTGGAGAACCCTCCGGCGGCTTCC 182
Db	61 TCCTCTGGAGAAATTTCACACAGTAGAGGCTGCCTTCATTGGAGAACCCTCCGGCGGCTTCC 120
OY	183 GGGAGAACCTCAATTATACCTACATCGCTCCTGTCTGTCTGTCAATCCCTACCGAG 242
Db	121 GGGAGAACCTCAATTATACCTACATCGGCTCCTGTCTGTCTGTCAATCCCTACCGAG 180
OY	243 ACCTACAGATCACAGCGCGGAGCATATGGAAAGCGCTACCGTGGTGTCAGATTCTPATGAAG 302
Db	181 ACCTACAGATCTACAGCGCGGAGCATATGGAAAGCTACCGTGGTGTCAGATTCTPATGAAG 240
OY	303 TACCACTCATTTGTTTGCAGTGGCTGCACATGTATACCGGGGCACTTGGTACTGAGCGTC 362
Db	241 TACCACTCATTTGTTTGCAGTGGCTGCACATGTATACCGGGGCACTTGGTACTGAGCGTC 300
OY	363 GGGACACAGGCAGTGATTTCTGGAGAGAGTGGGGCAGCAAGACAGAGCCACCACAAGA 422
Db	301 GGGACACAGGCAGTGATTTCTGGAGAGAGTGGGGCAGCAAGACAGAGCCACCACAAGA 360
OY	423 GACTCCTCAGTTCATGAGAGAGACTCGCCAGCCCCCTGAACGGGGGTGGCAGATGGCGAG 482
Db	361 GACTCCTCAGTTCATGAGAGAGACTCGCCAGCCCCCTGAACGGGGGTGGCAGATGGCGAG 420
OY	483 ACCGCTCTGTCAGAGCAACCCCGTGTAGAGGCTTTGGGAANTGCCAAGACTCTCCCGA 542
Db	421 ACCGCTCTGTCAGAGCAACCCCGTGTAGAGGCTTTGGGAANTGCCAAGACTCTCCCGA 480
OY	543 ACGATAACTCCAGCCGGTTTTGGAAGTACATGAGATGTGAGTTGACTTCAGAGGGGCCCC 602
Db	481 ACGATAACTCCAGCCGGTTTTGGAAGTACATGAGATGTGAGTTGACTTCAGAGGGGCCCC 540
OY	603 CCGTGGGAGGCCACATTTCTCAGTTACCTCTCGAAAAAGTCCCGGGGTGTGCACAAAATC 662
Db	541 CCGTGGGAGGCCACATTTCTCAGTTACCTCTCGAAAAAGTCCCGGGGTGTGCACAAAATC 600
OY	663 ACGGAGAGCGGGAATTCACAGTCTTTTACCACTACTGGAGGGGGCGGAGAGAGACTTC 722
Db	601 ACGGAGAGCGGGAATTCACAGTCTTTTACCACTACTGGAGGGGGCGGAGAGAGACTTC 660
OY	723 TCCGTCGGCTGGGCTTGGAAACGAACCCCACAAGTACTTGTACCTGTGAAGGGCCACT 782
Db	661 TCCGTCGGCTGGGCTTGGAAACGAACCCCACAAGTACTTGTACCTGTGAAGGGCCACT 720

OY	783	GTGCGAAGGCTCTCCATCATCAACGACAAAGTGTACTGGAGGTTATGAGAAAGCGCTGT	842
Db	721	GTGCGAAGGCTCTCCATCATCAACGACAAAGTGTACTGGAGGTTATGAGAAAGCGCTGT	780
OY	843	CCGTCATTGAGCTTCACTGAGAGTGAAGTGGAGACTTGGTCAGATCGAGGCGACGGCTC	902
Db	781	CCGTCATTGAGCTTCACTGAGAGTGAAGTGGAGACTTGGTCAGATCGAGGCGACGGCTC	840
OY	903	TACATCTGGGCAACATCCACTTTCGTGCTGACAGAGACAGCAATCCCGAGTTACTACTG	962
Db	841	TACATCTGGGCAACATCCACTTTCGTGCTGACAGAGACAGCAATCCCGAGTTACTACTG	900
OY	963	AGAACCAAGCTTCAAAATATCTGACCAAGGCTCTTGTGTGGAAGTACCAACTTAAAGGAAG	1022
Db	901	AGAACCAAGCTTCAAAATATCTGACCAAGGCTCTTGTGTGGAAGTACCAACTTAAAGGAAG	960
OY	1023	CCCTGACCCCAAGGAAGATCATGCGCCAAAGGGGAAGAGTCTCTGAGCCCACTGAACTTGG	1082
Db	961	CCCTGACCCCAAGGAAGATCATGCGCCAAAGGGGAAGAGTCTCTGAGCCCACTGAACTTGG	1020
OY	1083	AACAGGCGGCAATATGCAAGGAGTACGCTTGGCCAAAGGCTTGTACAGCGGCAATTCACCT	1142
Db	1021	AACAGGCGGCAATATGCAAGGAGTACGCTTGGCCAAAGGCTTGTACAGCGGCAATTCACCT	1080
OY	1143	GGCTGTGAGAAAGATCATATAGTCACTGGCCTTAAGAGCGCTGAGAGCCCAAGCTTGGC	1202
Db	1081	GGCTGTGAGAAAGATCATATAGTCACTGGCCTTAAGAGCGCTGAGAGCCCAAGCTTGGC	1140
OY	1203	GAAGCACCACGGGTTCTTGGGCTCCTGAGACATTTACGGCTTTTAAAGTGTTCAGCATAA	1262
Db	1141	GAAGCACCACGGGTTCTTGGGCTCCTGAGACATTTACGGCTTTTAAAGTGTTCAGCATAA	1200
OY	1263	GCTTCGAGCAAGTTCTGCAATCAACTGCAATAGAAAGTGTGAGAGGCTTTCATCAGCAGC	1322
Db	1201	GCTTCGAGCAAGTTCTGCAATCAACTGCAATAGAAAGTGTGAGAGGCTTTCATCAGCAGC	1260
OY	1323	TGACTCTCAAGTCGAGCAGAGGAATACGAGGCTGAGGGCATCGCGTGGGAACCTGTCC	1382
Db	1261	TGACTCTCAAGTCGAGCAGAGGAATACGAGGCTGAGGGCATCGCGTGGGAACCTGTCC	1320
OY	1383	AGTACTTCACACACACATATATCTGTGACTGGTAAAGAGAAGTTCAAGGGCATCATCT	1442
Db	1321	AGTACTTCACACACACATATATCTGTGACTGGTAAAGAGAAGTTCAAGGGCATCATCT	1380
OY	1443	CCATCTTGGATGAAGAGTGGCTCGTGCCTGGGGAGGCGCAGGACCTGACCTTTCGAGGA	1502
Db	1381	CCATCTTGGATGAAGAGTGGCTCGTGCCTGGGGAGGCGCAGGACCTGACCTTTCGAGGA	1440
OY	1503	AGTTTGAAGAGACACTGTCAACGCCACCACCCTCACTCTCTGACGACACAAGCTCGCTGACAGA	1562
Db	1441	AGTTTGAAGAGACACTGTCAACGCCACCACCCTCACTCTCTGACGACACAAGCTCGCTGACAGA	1500
OY	1563	AGACCAGGAATCCCTAGACCGAGGGAGTTCCGCTTTCGATTAATGCTGGAGAGGTGA	1622
Db	1501	AGACCAGGAATCCCTAGACCGAGGGAGTTCCGCTTTCGATTAATGCTGGAGAGGTGA	1560
OY	1623	CCTTACAGTGTACAGGGGTTCTTGGATTAAGAAACAATGACTCTCTTCGCGAACTCGAAGG	1682
Db	1561	CCTTACAGTGTACAGGGGTTCTTGGATTAAGAAACAATGACTCTCTTCGCGAACTCGAAGG	1620
OY	1683	AGACCATTGTGACGTCATATCAACCCCATCTATGCGCCCAAGTGTCTTGAACAAGATGAGCTCA	1742
Db	1621	AGACCATTGTGACGTCATATCAACCCCATCTATGCGCCCAAGTGTCTTGAACAAGATGAGCTCA	1680
OY	1743	GTGACAAAGAGCGGCCAG-GACGGTGGCCACCCAGTTCAAGATGAGCTCTGCAAGCTCG	1801
Db	1681	GTGACAAAGAGCGGCCAGAGACGGTGGCCACCCAGTTCAAGATGAGCTCTGCAAGCTCG	1740
OY	1802	TGGAGATCTCAGGTTCTTAAGAGCTGGCCATATTCGCGGTGATCAAGCCAAAGCAGCGCA	1861
Db	1741	TGGAGATCTCAGGTTCTTAAGAGCTGGCCATATATTCGCGGTGATCAAGCCAAAGCAGCGCA	1800
OY	1862	AGCAGCGGGGTGCTTGTGATGAGGTGCTCATCCGACATCAAGGTGAAGTACTCTGGGATCGA	1921

```
Db 1801 AGCAGCGGCGCTTGTGATGAGGTGCTATCCAGACATCAGTAACTGCGACTGA 1860
Qy 1922 TGGAGAAATCGCGCGTGGCCAGAGCGTGGTTCCTATCGTGAATATGAGCTTTCC 1981
Db 1861 TGGAGAAATCGCGCGTGGCCAGAGCGTGGTTCCTATCGTGAATATGAGCTTTCC 1920
Qy 1982 TGGAGAGTACAAAGTCTGTCGCGAGAGCATGCGCCATGTGGGCGAGCGGCCAGG 2041
Db 1921 TGGAGAGTACAAAGTCTGTCGCGAGAGCATGCGCCATGTGGGCGAGCGGCCAGG 1980
Qy 2042 ATGCTGTGCGCGTGTGGTGCAGACACTGCGCTACAAAGCCAGAGAGTACAAATGGCA 2101
Db 1981 ATGCTGTGCGCGTGTGGTGCAGACACTGCGCTACAAAGCCAGAGAGTACAAATGGCA 2040
Qy 2102 GAGTAAAGTCTGATCCGATTTCCCAAGACCTTATTTGCGCAGAGACCTCTGGAAG 2161
Db 2041 GAGTAAAGTCTGATCCGATTTCCCAAGACCTTATTTGCGCAGAGACCTCTGGAAG 2100
Qy 2162 TCCGCGCGGAGAGTCTAGCCAGAGATCCAGCGGCTGGAGGGCTTTCAATTGGGAG 2221
Db 2101 TCCGCGCGGAGAGTCTAGCCAGAGATCCAGCGGCTGGAGGGCTTTCAATTGGGAG 2160
Qy 2222 AGAATTTCTCGCGGTGAAGCATAGCCATCTGATCCAGTATGTTGGCGTGGCAGAC 2281
Db 2161 AGAATTTCTCGCGGTGAAGCATAGCCATCTGATCCAGTATGTTGGCGTGGCAGAC 2220
Qy 2282 TGGCGCGGAGAGGAGCCAGAGAGAAATGGGAGCGCCAGACCATCGTGCATCATCC 2341
Db 2221 TGGCGCGGAGAGGAGCCAGAGAGAAATGGGAGCGCCAGACCATCGTGCATCATCC 2280
Qy 2342 GTGGCTTCAATTTGCGCATTTCAACCCCGGTGGCTGAGAAATCCCTTCTTGGACACG 2401
Db 2281 GTGGCTTCAATTTGCGCATTTCAACCCCGGTGGCTGAGAAATCCCTTCTTGGACACG 2340
Qy 2402 TGGCGCGCTCAATTTGCTTAACTGAGGCGGCAACTGCGCGGAGATTTTGGACACT 2461
Db 2341 TGGCGCGCTCAATTTGCTTAACTGAGGCGGCAACTGCGCGGAGATTTTGGACACT 2400
Qy 2462 CTTGGCGGAGAGCCAGCTGCGCTGAGAGAGGCTCGAAGTCTGAGGAACTTGGACACT 2521
Db 2401 CTTGGCGGAGAGCCAGCTGCGCTGAGAGAGGCTCGAAGTCTGAGGAACTTGGAC 2460
Qy 2522 TGAAGAAATGCTGGAAGTACTGCGGAGACATCAACCCTGAGTGAAGAGAGAGCTGC 2581
Db 2461 TGAAGAAATGCTGGAAGTACTGCGGAGACATCAACCCTGAGTGAAGAGAGAGCTGC 2520
Qy 2582 AGCAAAAGGCGGTGCTAGTGAATTTTCAAGGCGAAGAGCAACTACCCCGAGAGTG 2641
Db 2521 AGCAAAAGGCGGTGCTAGTGAATTTTCAAGGCGAAGAGCAACTACCCCGAGAGTG 2580
Qy 2642 TCCCGAGACTTTCATTTAGCACAGGCTTGGCACAGAGAGATCAAGCCCGAGAGTG 2701
Db 2581 TCCCGAGACTTTCATTTAGCACAGGCTTGGCACAGAGAGATCAAGCCCGAGAGTG 2700
Qy 2702 AATCCTTGGGCTGTGAAGCCATCTCAGTATGCGGTCGCCGTGTAATATGAGCCGTAAG 2761
Db 2641 AATCCTTGGGCTGTGAAGCCATCTCAGTATGCGGTCGCCGTGTAATATGAGCCGTAAG 2700
Qy 2762 GTTACAAAGCTTCCCGGAGCTGCTGCTCAAGCCAGCCCACTGCTGCACTAGTACC 2821
Db 2701 GTTACAAAGCTTCCCGGAGCTGCTGCTCAAGCCAGCCCACTGCTGCACTAGTACC 2760
Qy 2822 ATGCTAAAGTCAAGAGAAATTTATGCAACCACTAACCCGGAATCTCTGCACTAGTACC 2881
Db 2761 ATGCTAAAGTCAAGAGAAATTTATGCAACCACTAACCCGGAATCTCTGCACTAGTACC 2820
Qy 2882 TGAGAGTAAAGCTTATTTGCTTACAGTGCAGGTGAAGAGCAACAGAGAGGAGATG 2941
Db 2821 TGAGAGTAAAGCTTATTTGCTTACAGTGCAGGTGAAGAGCAACAGAGAGGAGATG 2880
Qy 2942 TGGTGTGAGAGTGAATGATGATGAGACACTAACCAAGAGCGGCTCAGTGTGAGC 3001
|||||
```

```
Db 2881 TGGTGTGAGAGTGAATGATGATGAGACACTAACCAAGAGCGGCTCAGTGTGAGC 2940
Qy 3002 GCGTGAACAATATCAACATCAACAGGCGAGCATACGTTTGCAGGGGGTCCAGGAGG 3061
Db 2941 GCGTGAACAATATCAACATCAACAGGCGAGCATACGTTTGCAGGGGGTCCAGGAGG 3000
Qy 3062 AGCGATCATTTGACTTTCATCGGCGTCAAGAGCTTCATCAACCAAGGCTAAGATGGCC 3121
Db 3001 AGCGATCATTTGACTTTCATCGGCGTCAAGAGCTTCATCAACCAAGGCTAAGATGGCC 3060
Qy 3122 ACCTGCTGTGTGGTGGCGCCAGCGTGAATTTCTGGTGAAGAGCGTGGCGT 3172
Db 3061 ACCTGCTGTGTGGTGGCGCCAGCGCTGAATTTCTGGTGAAGAGCGTGGCGT 3111
```

```
RESULT 5
HSMYOIB
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
```

```
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
JOURNAL
```

FEATURES

source

## CDS

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="1/p3.2-p13.3"
/issue_type="kidney"
/dev_stage="adult"
66..3152
/codon_start=1
/product="myosin I beta"
/protein_id="CA61131.1"
/db_xref="GI:1926311"
/db_xref="SWISS-PROT:O00159"
/translation="MDSALTARDRVGVDQFLENTTSEAFTEIENRRRENIITY
IGPVIVSNPDRDLQITSROHMERVGVFEYVPHLTVADVADYVRLATERRDOAV
ISGSGGKTEATKILQFYAENCAPQAGVARNORLQNSPVEAEAGNKAIRNNS
RLEGKMDVDFEKGAPVGGHILSYLSESRVYNHBERNFHFTYOLLEGGEEPLR
LILGINIFANESNMAQYTEMQLYLRLSVEGSTLREALTHRKIAKGEELSPL
NLEQAVARALAKANTSTFTWLVGKINRSLADVESPSMSSTVVLGLIDYGEV
FOHNSFEORCINTCKEKLQDLFELPKSDEVEEAGETAMPVOYTFNNKICDLVEE
LHAYAEVYTVTFGLDNNDLFLNLEKEDTVKHHPRLTHKLADRTKSLRGSEFV
AGATRRKYEALQRYKSLCEPMPTWAGRPQGVAVLVNHLGYKPEEYKMRKTFET
REPRTLFAIEDALEVROSILATKTOAWRGFFHROKFLVKRSALCIGSMMRGITGR
KAARKMAQOTLRLIRGFLIRHAPRCPEBAFFLDHVRFSFLNLRPLPNVDTTR
PTPPALREASELRLBELCITKMMWAKTORSISPMKQOLQOKAVASEIFKKKNDIP
VFRLLSTRIGTDEISPRVLOALGSEPTIYAVPVYKDRGYKPPRSOLLILPNAVIA
VEDAKVORIDYANLTISVSLSDSIFVLHVRADRIKQGDVYIOSDHVLETJLTKTA
LSANVNSINTNOSTIFAGPGRDGTIFETGSELLITKAKNGLHAVARLNYR"
```

BASE COUNT

776 a 994 c 996 g 618 t

ORIGIN

Query Match 71.7%; Score 2626.2; DB 9; Length 3384;  
Best Local Similarity 87.1%; Pred. No. 0;  
Matches 2918; Conservative 0; Mismatches 428; Indels 5; Gaps 3;

OY 42 CGGCGCTGGCAGTACGCGGGTTCGAGTACGAGAGCGCCCTTGACGCCGAGACC 101  
DB 34 CGGCGCTGGCAGTACGCGGGTTCGAGTACGAGAGCGCCCTTGACGCCGAGACC 93  
OY 102 GGGTGGGGGTGAGAGCTTGTCTGCTGCGAGAAATTTTCCACGAGAGCGCTGCTCATTTG 161  
DB 94 GGGTGGGGGTGAGAGCTTGTCTGCTGCGAGAAATTTTCCACGAGAGCGCGCTTCTATCG 153  
OY 162 AGAACCTCGCGGCGGGTTCGCGGAGAACTCATTTATCTACATCGCTGCTGCTAG 221  
DB 154 AGAACCTCGCGGCGGGATTTGGGGAGAAATCTATCTACACCTACATTTGGCCCCGCTTCG 213  
OY 222 TCTCTGTAATCTCTTACCGAGACCTTACAGATCTACAGCGCGGACGATATGAAACCTTAC 281  
DB 214 TCTCTGTAATCTCTTACCGGAGACCTTACAGATCTACAGCGCGGACGATATGAAACCTTAC 273  
OY 282 GGGTGGGTGAGTTCATGAGTACGACCCATTTGTTGCAATGGCTGACACTGATATACC 341  
DB 274 GTGGGCTGAGCTTCTATGAGTACGCCCCCTCACCTGTTGCGTGGGACACTGCTGATACC 333  
OY 342 GGGCACTTCTGTAAGGCTCGGAGACAGGAGTATGATTTCTGGAGAGTGGGGCAG 401  
DB 334 GAGCACTGCGGACAGGAGTGGGAGACAGGCTGATGATCTCTGGGAGAGCGGGGCGAG 393  
OY 402 GCAGACAGAGGCCACCAAGAGACTGCTCAGTTCTATGACAGAGACTTGGCCAGCCCTG 461  
DB 394 GCAGACAGGAGGCCACCAAGAGCTGCTCAGTTCTATGACAGAGACTTGGCCAGCCCTG 453  
OY 462 AACGGGTGGGCGGAGGAGAGCGCCCTTTCGAGAGCAACCCCGTTTGAAGGCTTGG 521  
DB 454 AACGGGTGGGCGGAGGAGAGCGCCCTTTCGAGAGCAACCCCGTTTGAAGGCTTGG 513  
OY 522 GGAATGCGCAAGACTCTCGGCAAGATTAATCTCAGCCGGTTTGAAGTACATGATGTC 581  
DB 514 GGAATGCGCAAGACTCTCGGCAAGATTAATCTCAGCAGTTCGGGAAAGTACATGATGTC 573  
OY 582 ACTTGAATCTCAAGGCTGCGCCGCTGGGAGGCGCAATCTCAGTTAATCTCTGGAAAGT 641  
DB 574 AGTTGACTTCAAGGCTGCGCCGCTGGGAGGCGCAATCTCAGTTAATCTCTGGAAAGT 633  
OY 642 CCGGGGTGGTGCACCAAAATCAAGAGAGCGGAACTTCCACGCTTTTACGAGTACTGG 701  
DB 634 CAGGAGTGGTGCACCAAAATCAAGAGAGCGGAACTTCCACATCTTTCACAGAGTCTGG 693  
OY 702 AGGGGGGCGAGAGAGAGTCTCGCTCGGCTGGGCTTGGAGAGAAACCCCGAGAGTACT 761  
DB 694 AGGGGGGCGAGAGAGAGTCTCGGCTGGGCTTGGAGAGAAACCCCGAGAGTACT 753  
OY 762 TGTACCTGCTGAAGGCGCAAGTGTGCCAAGGCTCTCTCATCAACGACAAAGTACTGGA 821  
DB 754 TGTACCTGCTGAAGGCGCAAGTGTGCCAAGGCTCTCTCATCAACGACAAAGTACTGGA 813  
OY 822 AGCTATGAGGAAGGCGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 881  
DB 814 AGCTATGAGGAAGGCGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 873  
OY 882 TCAGCATGCTGAGCGAGCTCTACATCTGGGCAAGATCATCTCTCTCTCTCTCTCTCTCTCT 941  
DB 874 TAAAGTATGCTGGCCAGCGCTTCAATTTTGGGCAACATCTCTCTCTCTCTCTCTCTCTCT 933  
OY 942 GCAATGCGCCAGGTTACTACTGAGAAACAGCTCAAAATATCTGACAGGCTCTTGTGTGG 1001  
DB 934 GCAATGCGCCAGGTTACTACTGAGAAACAGCTCAAAATATCTGACAGGCTCTTGTGTGG 993  
OY 1002 AAGGTACAAACATTTAGGAGAGCGCTGACCCACAGAGAAATCTCCCAAGGGGAGAGAGC 1061  
DB 994 AAGGTACAAACATTTAGGAGAGCGCTGACCCACAGAGAAATCTCCCAAGGGGAGAGAGC 1053

OY 1062 TCCTGAGCCCACTGAACCTTGAACAGCGCATATGCAAGGATGCGCTTGCCAAAGCTG 1121  
DB 1054 TCCTGAGCCCACTGAACCTTGAACAGCGCATATGCAAGGATGCGCTTGCCAAAGCTG 1113  
OY 1122 TGTACAGCCGGAATTCACCTGCTGTGTGAGAAAGATCAATAGCTACTGCTCTAAGG 1181  
DB 1114 TGTACAGCCGGAATTCACCTGCTGTGTGAGAAAGATCAATAGCTACTGCTCTAAGG 1173  
OY 1182 ACGCTGAGAGCCCACTGCTGAGAGAGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1241  
DB 1174 ACGTGAAGAGCCCACTGCTGAGAGAGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1233  
OY 1242 TTGAAGTGTTCAGCATTAACAGCTTTCGAGAGTTCGATCACTCACTCACTCACTCACTCA 1301  
DB 1234 TCGAAGTGTTCAGCATTAACAGCTTTCGAGAGTTCGATCACTCACTCACTCACTCACTCA 1293  
OY 1302 TCGAGAGCTTTCATGAGCTGACCTCAAGTGGAGAGAGAGAAATACGAGCTGAGG 1361  
DB 1294 TCGAGAGCTTTCATGAGCTGACCTCAAGTGGAGAGAGAGAAATACGAGCTGAGG 1353  
OY 1362 GCATGCGGTGGGAGCCCTGCTGAGTACCAAGATGATGATGATGATGATGATGATGATG 1421  
DB 1354 GCATGCGGTGGGAGCCCTGCTGAGTACCAAGATGATGATGATGATGATGATGATGATG 1413  
OY 1422 AGAAGTTCAAGGCGCATCATCTCCATCTTGGATGAAGAGTGCCTGCTCTGGGAGGCCA 1481  
DB 1414 AGAAGTTCAAGGCGCATCATCTCCATCTTGGATGAAGAGTGCCTGCTCTGGGAGGCCA 1473  
OY 1482 CGGACCTGACCTTTCGAGAGAGTGGAGAGACCTGTAAGCCCACTCACTCTCTGA 1541  
DB 1474 CAGACCTGACCTTTCGAGAGAGTGGAGAGACCTGTAAGCCCACTCACTCTCTGA 1533  
OY 1542 CGCACAGCTGCTGCTGAG 1601  
DB 1534 CGCACAGCTGCTGCTGAG 1593  
OY 1602 TGCATTTCTGAGAGAGTGCATCTACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1661  
DB 1594 TGCATTTCTGAGAGAGTGCATCTACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1653  
OY 1662 TCTCTTCCGGAACCTGAG 1721  
DB 1654 TCTCTTCCGGAACCTGAG 1713  
OY 1722 GCTTGAAGAGAGTGAAGTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1780  
DB 1714 GCTTGAAGAGAGTGAAGTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1773  
OY 1781 AGATGAGCTCTGCAAGCTGCTGAGAGATCTGAGGTCTAAGGAGCTGCTATATCCGCT 1840  
DB 1774 AGATGAGCTCTGCAAGCTGCTGAGAGATCTGAGGTCTAAGGAGCTGCTATATCCGCT 1833  
OY 1841 GCATCAAGCCAAACGAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1900  
DB 1834 GCATCAAGCCAAACGAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1893  
OY 1901 AGGTGAAGTACGCGGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1960  
DB 1894 AGGTGAAGTACGCGGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1953  
OY 1961 GTGCAAAATATGAGGCTTCTGCAAGAGTACAAAGTCACTGTGCCCCAGAGAGATGAGCCA 2020  
DB 1954 GCGGCAAAATATGAGGCTTCTGCAAGAGTACAAAGTCACTGTGCCCCAGAGAGATGAGCCA 2013  
OY 2021 TGTGGGAGAGAGCGCCCAAGAGATGCTGCGGTGTGTGAGACACCTGCGCTACAAAC 2080  
DB 2014 CGTGGGAGAGAGCGCCCAAGAGATGCTGCGGTGTGTGAGACACCTGCGCTACAAAC 2073  
OY 2081 CAGAAGAGTCAAAATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2140  
DB 2074 CAGAAGAGTCAAAATAGGAG 2133









Db	2170	CACAGAGAGCGCTTGGACATCCGACGACAGACAGCTGGCCACCAAGATCCAGCCACCTG	2229
Qy	2202	GAGGGCTTTCATTTGGCGACAGAAATTTCTCGGGGTAAAGCGATTCAGCCATATGTAATCA	2261
Db	2230	GAGGGCTTTCATTTGGCGACAGAAATTTCTCGGGGTAAAGCGATTCAGCCATATGTAATCA	2261
Qy	2262	GTCATGTCGCTGGCGACACAGTGGCGGAGAAAGGACAGCCCAAGAGATGTCGATCCCA	2289
Db	2290	GTCATGTCGCTGGCGACACAGTGGCGGAGAAAGGACAGCCCAAGAGATGTCGATCCCA	2321
Qy	2322	GACATCCGCTGACTCATTCGCTGGCTTCATTTTGGCCATTCACCCCGTGGCCCTGAGAA	2349
Db	2350	GACATCCGCTGACTCATTCGCTGGCTTCATTTTGGCCATTCACCCCGTGGCCCTGAGAA	2381
Qy	2382	TGCTTCTTCTTGGACACAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	2409
Db	2410	TGCTTCTTCTTGGACACAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	2441
Qy	2442	CCGGATTTTCTTGGACACAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	2469
Db	2470	CCGGATTTTCTTGGACACAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	2501
Qy	2502	ACTGCTAGGGAGACAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	2529
Db	2530	ACTGCTAGGGAGACAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	2561
Qy	2562	TGAGTGGAGACAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	2589
Db	2590	TGAGTGGAGACAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	2621
Qy	2622	GGACAACTTACCCCGACAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	2649
Db	2650	GGACAACTTACCCCGACAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	2681
Qy	2682	GATCAGCCCGACAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	2709
Db	2710	GATCAGCCCGACAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	2741
Qy	2742	GATCAGCCCGACAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	2769
Db	2770	GATCAGCCCGACAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	2801
Qy	2802	TGCTTCTTCTTGGACACAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	2829
Db	2830	TGCTTCTTCTTGGACACAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	2861
Qy	2862	CGGATTCCTGTCACAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	2889
Db	2890	CGGATTCCTGTCACAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	2921
Qy	2922	CAGCAAGAGAGAGAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	2949
Db	2950	CAGCAAGAGAGAGAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	2981
Qy	2982	GACGAGCCCTGACAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	3009
Db	3010	GACGAGCCCTGACAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	3041
Qy	3042	TGCAGGGGTCAGAGAGAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	3069
Db	3070	TGCAGGGGTCAGAGAGAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	3101
Qy	3102	CAGCAAGAGAGAGAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	3129
Db	3130	CAGCAAGAGAGAGAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	3161
Qy	3162	AAAGG-TCGCGTGGAGAGAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	3189
Db	3190	AAAGG-TCGCGTGGAGAGAGTGGCGGCTCAATTTTGGCTTACCTTACCTGCGCCCTGAGAA	3220
Qy	3221	GACTTACCAAAAGTCAAAAGAGGATTCATGACAGACCCCTCAAAAGCCACGCT	3248
Db	3248	GACTTACCAAAAGTCAAAAGAGGATTCATGACAGACCCCTCAAAAGCCACGCT	3280

Db	3249	CAGTACCAAAAGTCAAAAGAGGATTCATGACAGACCCCTCAAAAGCCACGCT	3280
Qy	3281	GCAAACTCTGCTTCTTGGCGACAGAAATTTCTCGGGGTAAAGCGATTCAGCCATATGTAATCA	3306
Db	3307	AAACTCGCCCTCTCTCT--CGGCTTCTTCAGAGGGGGCGGACAGGAGAAAGCTGCGCC	3340
Qy	3341	ATGAGTGGGCGGACAGGCGGCGGACACCAATTAAGCAAGAGGATTTTAACTGAGG--TTTCTCTGA	3423
Db	3365	AGGAGT--GGCAGCGCGGCGGACCAAGCAATTAAGCAAGAGGATTTTAACTGAGG--TTTCTCTGA	3458
Qy	3401	G--CTCTGCTGATGCAAAATATCTTAAGCAAGAGGATTTTAACTGAGG--TTTCTCTGA	3483
Db	3424	GCCCGACGCGCAATGCAAAATTTTAAAGAGGAGGAACTTTGCTGAGGTTTTTCTGA	3483
Qy	3459	GATTTTTT 3466	
Db	3484	GATTTTTT 3491	

RESULT 7	BTMOIBA	3148 bp	mRNA	Linear	MM 10-FEB-1999
LOCUS	DEFINITION	B. taurus myosin IB mRNA, complete cds.			
ACCESSION	222852				
VERSION	222852.1	GI:397402			
KEYWORDS	myosin IB.				
SOURCE	ORGANISM	Bos taurus			
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.			
REFERENCE	AUTHORS	Reizes, O., Barryko, B., Li, C., Sudhof, T.C. and Albanesi, J.P.			
	TITLE	Domain structure of a mammalian myosin I beta			
	JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 91 (14), 6349-6353 (1994)			
	MEDLINE	94294379			
	PUBMED	8022785			
REFERENCE	AUTHORS	Reizes, O.			
	TITLE	Direct Submission			
	JOURNAL	Pharmacology, 5323 Harry Hines Boulevard, Dallas, Texas, USA, 75209			
COMMENT	FEATURES	On Sep 3, 1993 this sequence version replaced gi:3115101.			
source		1. 3148			
		/organism="Bos taurus"			
		/db_xref="taxon:9913"			



Db 1921 CGCAATACGAGGCTTCTCGAGAGGTACAGTCACTGTGCCCCAGAGACATGCGCCACG 1980  
 QY 2023 TGGCAGAGAGGCGCCAGAGTGTGTGCGCCGTGTGTGACAGACACTGCGCTACAGAGCA 2082  
 Db 1981 TGCACAGAGAGCTCCAGAGATGGGGGTGACTGTCTGTGTAGGACCTGGGCTACAGCCG 2040  
 QY 2083 GAAGAGTCCAAATATGGCGAGAGTCAAGATCTTCACTTCCGATTCGCCAAGACCTTATTTGCC 2142  
 Db 2041 GAAGAGTCCAAATATGGCGAGAGTCAAGATCTTCACTTCCGATTCGCCAAGACCTTATTTGCC 2142  
 QY 2143 ACAGAGACTTCCCTGGAGAGTGGCGGAGAGTGTAGCAGCAAGATCCAGGCGGCTTGG 2202  
 Db 2101 ACAGAGAGCTCTTGGAGATCCGACAGAGCTGTGCGCAGGAATCCAGGCGCACTGG 2160  
 QY 2203 AGGCGCTTCAATTTGGCGAGAGAAATTTCTCCGGGTGAAGGAGTACCCATCTGTATCCAG 2262  
 Db 2161 AGGCGCTTCACTTCCGCGGAGAAATTTCTCCGGGTGAAGGAGTACCCATCTGTATCCAG 2262  
 QY 2263 TCATGTGTGCGCTGCGACACTGTGGCGGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAG 2322  
 Db 2221 TCGTGTGTGGAGAGACGTGGCGCGGAGAGAGGCGGCGCAAGAGAGAGAGAGAGAGAGAG 2322  
 QY 2323 ACCATCCGCTGACTCATCCGTGTGCTTCAATTTTGGCCATTCACCCGCTGCTGAGAT 2382  
 Db 2281 ACCATCCGCGGCTCATCCAGGGCTTCACTTGTGCGCATGTGCGCCGCTTGTGAGAT 2340  
 QY 2383 GCTTCTCTTGGACACAGCTGTGGCGCTTCAATTTTGGCTTCACTTGTGCGCATGTGCGCC 2442  
 Db 2341 GCTTCTCTTGGACACAGCTGTGGCGCTTCAATTTTGGCTTCACTTGTGCGCATGTGCGCC 2442  
 QY 2443 CGGATGTTTGTGGACACCTGTGCGCCACACCCCGGCTGTGAGAGAGAGAGAGAGAGAGAG 2400  
 Db 2401 CGGATGTTTGTGGACACCTGTGCGCCACACCCCGGCTGTGAGAGAGAGAGAGAGAGAGAG 2400  
 QY 2503 CTGTACCGGAACTGTGATGAGAAACATGTGTGGAAGTACTGCGGACATCAAGCCCT 2562  
 Db 2461 CTGTGTGGGAACTGTGCGGAGAAACATGTGTGGAAGTACTGCGGACATCAAGCCCT 2520  
 QY 2563 GAGTGGAG 2622  
 Db 2521 GAGTGGAG 2622  
 QY 2623 GAGTGGAG 2682  
 Db 2581 GAGTGGAG 2682  
 QY 2683 ATACAGCCAG 2640  
 Db 2641 ATACAGCCAG 2640  
 QY 2743 GTAAATATGAG 2700  
 Db 2701 GTAAATATGAG 2700  
 QY 2803 GCTGTGTGCTATTTGTGAG 2760  
 Db 2761 GCTGTGTGCTATTTGTGAG 2760  
 QY 2863 GGAATCTCTGCACTAG 2820  
 Db 2821 GGAATCTCTGCACTAG 2820  
 QY 2923 AACAG 2880  
 Db 2881 AACAG 2880  
 QY 2983 AGGCGCTTCAATTTGGCGAGAGAAATTTCTCCGGGTGAAGGAGTACCCATCTGTATCCAG 2940  
 Db 2941 AGGCGCTTCACTTCCGCGGAGAAATTTCTCCGGGTGAAGGAGTACCCATCTGTATCCAG 2940  
 QY 3043 GCAGGGGTTCAG 3000  
 Db 3001 GCAGGGGTTCAG 3000

QY 3103 ACCAGGCTTCAAGATGGCCACCTGCTGTGTGTGGCGCCAGCGCTGATTCGGGTATGCA 3162  
 Db 3061 ACCAGGCTTCAAGATGGCCACCTGCTGTGTGTGGCGCCAGCGCTGATTCGGGTATGCA 3162  
 QY 3163 AGCG-TGGGTGTGAG 3189  
 Db 3121 AGCGCCCGAG 3120

RESULT 8  
 LOCUS MMTB  
 DEFINITION M.musculus mRNA for myosin-I beta.  
 ACCESSION X99638.1 GI:1924960  
 VERSION X99638.1 GI:1924960  
 KEYWORDS myosin-I beta.  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 Reference  
 1 (bases 1 to 2568)  
 Crozet, F.  
 Direct Submission  
 Submitted (30-JUL-1996) F. Crozet, Inst. Pasteur, Unite de Genetique Moleculaire Humaine, Institut PASTEUR, 25 rue du Docteur Roux, F- 75724, PARIS, cedex 15, FRANCE  
 2 (bases 1 to 2568)  
 Crozet, F., Amroun, A.E., Blanchard, S., Lenoir, M., Ripoll, C., Vago, P., Hamel, C., Fizames, C., Levl-Acobas, F., Depetris, D., Mettel, M.G., Weil, D., Pujol, R. and Pell, C.  
 Cloning of the genes encoding two murine and human cochlear unconventional type I myosins  
 Genomics 40 (2), 332-341 (1997)  
 9119401

FEATURES  
 source  
 1..2568  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /strain="BALB/c"  
 /db\_xref="taxon:10090"  
 /tissue\_type="cochlea"  
 /dev\_stage="adult"  
 147..>2568  
 /codon\_start=1  
 /product="myosin-I beta"  
 /protein\_id="CA67956.1"  
 /db\_xref="GI:1924961"  
 /db\_xref="MGD:MGI:106612"  
 /translation="MESLJARDRVGVDDFVLENFSEAFLENLRGFEENLTYT  
 IGSFGAGCTEATKRLQFYAEATGAPVGHILSLYLEKSRVHONHGRNPFVYOLLEGGEBETLR  
 RLGEENRPOSYLIVKGGCAVENSINDKSDMKVRKALSLVIDFEDPVEDLSIVSV  
 NLEDAVYARDALAKAVYSRTPTMLVLRKINSLSADSPSRSTTYVGLIDYGFEE  
 FQNSFEQFCINAKENKIDOLFTEBLTKSDEEYFAEIGAMPVYQITNNKIIDLVEE  
 LHVAGEVTVSVTEGLDKNDLLENLKJETSMSNPIMACGLADOKTRKSLDRGFVA  
 TQPKWSLLOULEILRSKRYATRCIPNDAKOPGPPDV LARHOVATYGLIMENLAVR  
 RFPTLTLPATDLSLEVRKSLATKIOAMRGFHMOKTLRVKRSALICIQSMWGTGTR  
 KAARKVAADTIRILRGLIRHNSPRCPENAFILDHVRSFLNLGRQLPRNVLDTSN  
 PTPPPALRD"

BASE COUNT 593 a 710 c 751 g 514 t

Query Match  
 Best Local Similarity 66.08; Score 2417.6; DB 10; Length 2568;  
 Matches 2437; Conservative 0; Mismatches 14; Indels 1; Gaps 1;  
 QY 42 CGGCCCTGGGCACTGACGGGCTTGTGAGTGCATGAGAGAGCGCTTGTACTGCCGAGACC 101

|||||  
 Db 115 CGGCGCTGGCGAGTACGCGGGTTCCAGTGCACCATGGAGAGCGCTTGACTGCCGAGACC 174  
 Oy 102 GGGTAGGGGTGAGGACTTTGTCGCTGAGNAATTTACCGCTGAGCGTGCCTTCATTG 161  
 Db 175 GGGTAGGGGTGAGGACTTTGTCGCTGAGNAATTTACCGCTGAGCGTGCCTTCATTG 234  
 Oy 162 AGAACCTCGCGGGGGGTTCCGGGAGAACCTTATTAACCTGCTGCTGCTAG 221  
 Db 235 AGAACCTCGCGGGGGGGTTCCGGGAGAACCTTATTAACCTGCTGCTGCTAG 294  
 Oy 222 TCTCTGTAATCCCTACCGAGACTTACAGATCTACAGCCGAGCATATGGAAGCTTACC 281  
 Db 295 TCTCTGTAATCCCTACCGAGACTTACAGATCTACAGAGGAAGCATATGGAAGCTTACC 354  
 Oy 282 GGGGTGTCAGTTTCTATGAATGACACCTCATTTGTTGTCATGCTGACATGTAATACC 341  
 Db 355 GTGGTGTCAATTTCTATGAATGACACCTCATTTGTTGTCATGCTGACATGTAATACC 414  
 Oy 342 GGGCACTTCGTACTGAGCGTGGGACACAGGAGATGATGATTTCTGAGAGAGTGGGGCAG 401  
 Db 415 GGGCACTTCGTACTGAGCGTGGGACACAGGAGATGATGATTTCTGAGAGAGTGGGGCAG 474  
 Oy 402 GCAACACAGAGGCGCACCAAGAGACTGCTCAGTTCTATGCAAGAGACTTCCAGCCCTG 461  
 Db 475 GCMAACAGAGGCGCACCAAGAGACTGCTCAGTTCTATGCAAGAGACTTCCAGCCCTG 534  
 Oy 462 AACGGGGGCGCGAGTGGGAGACGCGCTGTTGACAGAACCCCGTTGAGAGGCGCTTGG 521  
 Db 535 AACGGGGTGGCGCAGTGGGAGACGCGCTGTTGACAGAACCCCGTTGAGAGGCGCTTGG 594  
 Oy 522 GGAATGCCAAGACTCTCGGCAACGATTAATCTCAGCGCGTTTGAAAGATACATGATGTC 581  
 Db 595 GGAATGCCAAGACTCTCGGCAACGATTAATCTCAGCGCGTTTGAAAGATACATGATGTC 654  
 Oy 582 AGTTTGACTTCAAGGGTGGCCCCGTTGGAGGCCACATTTCTCAGTTAATCTCTGGAAAGT 641  
 Db 655 AGTTTGACTTCAAGGGTGGCCCCGTTGGAGGCCACATTTCTCAGTTAATCTCTGGAAAGT 714  
 Oy 642 CCCGGGTGGTGCACCAAAATACGAGAGCGAGACTTCCAGCTTTTTCAGACTATCG 701  
 Db 715 CACGAGTGGTGCACCAAAATACGAGAGCGAGACTTCCAGCTTTTTCAGACTATCG 774  
 Oy 702 AGGGGGGCGAGGAGAGACTCTCGCTCGGCTGGGCTTGGAGCGAACCCCGAGACTACT 761  
 Db 775 AGGGGGGCGAGGAGAGACTCTCGCTCGGCTGGGCTTGGAGCGAACCCCGAGACTACT 834  
 Oy 762 TGTACTGTGTAAGGGCCAGTGTGCCAAGGTCTCTCCATCAACGACAAGAGTACTGGA 821  
 Db 835 TGTACTGTGTAAGGGCCAGTGTGCCAAGGTCTCTCCATCAACGACAAGAGTACTGGA 894  
 Oy 822 AGTTATGGAAGGCGCTGCTCGTCAATGACTTCACTAGAGATGAAGGAGGACTTGC 881  
 Db 895 AGTTATGGAAGGCGCTGCTCGTCAATGACTTCACTAGAGATGAAGGAGGACTTGC 954  
 Oy 882 TCAGCATCGTGGCCAGCGTCTTACATCTGAGGCAACATCCACTTGTGCTGAGAGGACA 941  
 Db 955 TCAGCATCGTGGCCAGCGTCTTACATCTGAGGCAACATCCACTTGTGCTGAGAGGACA 1014  
 Oy 942 GCAATGCGCCAGGTACTGTAAGAACCAAGTCAATATCTGACAGGCTCTTGGTGTGG 1001  
 Db 1015 GCAATGCGCCAGGTACTGTAAGAACCAAGTCAATATCTGACAGGCTCTTGGTGTGG 1074  
 Oy 1002 AAGGTACAACTTTAGGGAAGCGCTGACCCACAGGAAGTCAATCGGCAAGGGGGAAGAGC 1061  
 Db 1075 AAGGTACAACTTTAGGGAAGCGCTGACCCACAGGAAGTCAATCGGCAAGGGGGAAGAGC 1134  
 Oy 1062 TCCTGAGCCCACTGAACCTTGAACAGGCGGATATGCAAGGATGGCTTGGCAAGGCTG 1121  
 Db 1135 TCCTGAGCCCACTGAACCTTGAACAGGCGGATATGCAAGGATGGCTTGGCAAGGCTG 1194  
 Oy 1122 TGTACAGCGGACATTTACCTGGCTGGTGAAGATCAATAGCTACTGGCTCTTAAG 1181  
 |||||

Db 1195 TGTACAGCGGACATTTACCTGGCTGGTGAAGAAATCAATAGTCACTGGCCCTTAAG 1254  
 Oy 1182 ACCTGAGAGCGCCAGCTGGCGAAGCACCGGTTTGGGCGCTGGACATTTACGCT 1241  
 Db 1255 ACCTGAGAGCGCCAGCTGGCGAAGCACCGGTTTGGGCGCTGGACATTTACGCT 1314  
 Oy 1242 TTGAAGTGTTCAGCATTAACAGCTTCGAGCAGTTCATCAACTCTGCAATGAGAAC 1301  
 Db 1315 TTGAAGTGTTCAGCATTAACAGCTTCGAGCAGTTCATCAACTCTGCAATGAGAAC 1374  
 Oy 1302 TGCAGCAGCTTTTCATCGAGCTGACTCTAAGTCGAGACAGAGAAATACGAGCTGAG 1361  
 Db 1375 TGCAGCAGCTTTTCATCGAGCTGACTCTAAGTCGAGACAGAGAAATACGAGCTGAG 1434  
 Oy 1362 GCATCGCTGGGAGACCTGACAGTACTTGAACACAAAGTCAATCTGACCTGGTGAAG 1421  
 Db 1435 GCATCGCTGGGAGACCTGACAGTACTTGAACACAAAGTCAATCTGACCTGGTGAAG 1494  
 Oy 1422 AGAAGTTCAAGGGCATCATCTCCATCTGATGAAGATGTCCTGCTGGGAGGCA 1481  
 Db 1495 AGAAGTTCAAGGGCATCATCTCCATCTGATGAAGATGTCCTGCTGGGAGGCA 1554  
 Oy 1482 CGGACCTGACCTTTCTGAGAAAGTTGGAGGACACTGCAAGCCCACTCACTTCTGA 1541  
 Db 1555 CGGACCTGACCTTTCTGAGAAAGTTGGAGGACACTGCAAGCCCACTCACTTCTGA 1614  
 Oy 1542 CGCACAGCTCGCTGACACAGAAAGACAGGAAATCCCTACACGAGGGAGTTCCGCTTC 1601  
 Db 1615 CGCACAGCTCGCTGACACAGAAAGACAGGAAATCCCTACACGAGGGAGTTCCGCTTC 1674  
 Oy 1602 TGCATTTATCGTGAGAGGTGACCTTACAGTGTGACTGGGTTTCTGATTAATAAATGACC 1661  
 Db 1675 TGCATTTATCGTGAGAGGTGACCTTACAGTGTGACTGGGTTTCTGATTAATAAATGACC 1734  
 Oy 1662 TCCCTCTCCGGAACCTGGAAGGAGACCATGTGACGCTCAATGAACCCCATCATGCCCCA 1721  
 Db 1735 TCCCTCTCCGGAACCTGGAAGGAGACCATGTGACGCTCAATGAACCCCATCATGCCCCA 1794  
 Oy 1722 GCTTTGACAAGAGTGAAGTCAAGTGAACAAGAGGCGGACAG-GACGGTGGGCGACCCAGTTCA 1780  
 Db 1795 GCTTTGACAAGAGTGAAGTCAAGTGAACAAGAGGCGGACAGGAGTGGCCACCCAGTTCA 1854  
 Oy 1781 AGATGAGCCTCTGACAGCTGCTGAGAGATCCTGAGGCTTAAGGACCTGCTATATCCGT 1840  
 Db 1855 AGATGAGCCTCTGACAGCTGCTGAGAGATCCTGAGGCTTAAGGAGCCTGCTATATCCGT 1914  
 Oy 1841 GCATCAAGCCAAACGACGCCAAGACCGCGGCTGCTTTGATGAGTGTCTATCCGACATC 1900  
 Db 1915 GCATCAAGCCAAACGACGCCAAGACCGCGGCTGCTTTGATGAGTGTCTATCCGACATC 1974  
 Oy 1901 AGGTGAAGTACCTGGGAGCTGATGGAATCTGCGCTGGCGACAGAGCTGGCTTGGCTATC 1960  
 Db 1975 AGGTGAAGTACCTGGGAGCTGATGGAATCTGCGCTGGCGACAGAGCTGGCTTGGCTATC 2034  
 Oy 1961 GTGCGAAATATGAGGCTTTCTCTGACAGAGGTACAAGTCACTGTGCCAGAGATGGCCCA 2020  
 Db 2035 GTGCGAAATATGAGGCTTTCTCTGACAGAGGTACAAGTCACTGTGCCAGAGATGGCCCA 2094  
 Oy 2021 TGTGGGCGAGAGCGGCCCGAGAGATGTTGTCGCTGTTGGTCAACACCTCGGTACAAGC 2080  
 Db 2095 TGTGGGCGAGAGCGGCCCGAGAGATGTTGTCGCTGTTGGTCAACACCTCGGTACAAGC 2154  
 Oy 2081 CAGAAAGTACAAAATGGGCGAGGACTTAAGATCTTCAATCGATGCCCAAGAGCTTATTTG 2140  
 Db 2155 CAGAAAGTACAAAATGGGCGAGGACTTAAGATCTTCAATCGATGCCCAAGAGCTTATTTG 2214  
 Oy 2141 CCACAGAGACTTCCCTGGAAGTCCGGCGGACAGTCTAGCCACCAAGATCCAGCGCGCT 2200  
 Db 2215 CCACAGAGACTTCCCTGGAAGTCCGGCGGAGAGTCTAGCCACCAAGATCCAGCGCGCT 2274  
 Oy 2201 GGAAGGCTTTTCAATTTGGCGACAGAAATTTCTCGGGGTGAAGCGATCAGCATCTGTATCC 2260  
 Db 2275 GGAAGGCTTTTCAATTTGGCGACAGAAATTTCTCGGGGTGAAGCGATCAGCATCTGTATCC 2334  
 |||||

QY 2261 AGTCATGTCGCGTGGACACCTGGCCGAGAGAGCAGCAGAGAAAGTGGCCAGCCC 2320  
 Db 2335 AGTCATGTCGCGTGGACACCTGGCCGAGAGAGCAGCAGAGAAAGTGGCCAGCCC 2320  
 QY 2321 AGCAGATCGCTGACATCCGTCCTTCATTTTGGCCCATTCACCCGCTCCCTGAGA 2380  
 Db 2395 AACACATCCGTCGACATCCGTCCTTCATTTTGGCCCATTCACCCGCTCCCTGAGA 2380  
 QY 2381 ATGCTTCCTTCTGGACACCTGCGCCCTCATTTTGGTCCTAACCTGGCGGCACTGC 2454  
 Db 2455 ATGCTTCCTTCTGGACACCTGCGCCCTCATTTTGGTCCTAACCTGGCGGCACTGC 2454  
 QY 2441 CCCGGAATGTCCTGGACACCTGCGCCCTCATTTTGGTCCTAACCTGGCGGCACTGC 2514  
 Db 2515 CCCGGAATGTCCTGGACACCTGCGCCCTCATTTTGGTCCTAACCTGGCGGCACTGC 2514

RESULT 9  
 LOCUS RCU14382 3288 bp mRNA linear VRT 18-APR-1995  
 DEFINITION Rana catesbeiana myosin I beta mRNA, complete cds.  
 ACCESSION U14382  
 VERSION U14382.1 GI:603691  
 KEYWORDS  
 SOURCE  
 ORGANISM Rana catesbeiana.  
 Rana catesbeiana.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumleostomi;  
 Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Ranidae; Rana;  
 Aquarana.  
 1 (bases 1 to 3288)  
 SOLC,C.F., Deffler,B.H., Duyk,G.M. and Corey,D.P.  
 Molecular cloning of myosins from the bullfrog sacculus macula: A  
 candidate for the hair-cell adaptation motor  
 Aud. Neurosci. 1, 63-75 (1994)  
 2 (bases 1 to 3288)  
 Corey,D.P.  
 Direct Submission  
 Submitted (05-SEP-1994) David P. Corey, Massachusetts  
 Hospital, Boston, MA 02114, USA  
 Location/Qualifiers  
 1. 3288  
 /organism="Rana catesbeiana"  
 /db\_xref="taxon:8400"  
 /tissue\_type="kidney"  
 /dev\_stage="adult"  
 68. 3154  
 /codon\_start=1  
 /product="myosin I beta"  
 /protein\_id="AA65091.1"  
 /db\_xref="GI:603692"  
 /translation="MESALTPARDRVQDPEVLLNNTSEAFIENLRKREKULITY  
 IGSVLAVSNPYKLEIYSKOHMERIRGVSEVSPHLYALADNSYRSLRTERKULITY  
 ISGSGAKTEASKKTILOYTAVTCPSVQSVETVDRLLQSPVLEPFGAKTLDNOCIL  
 SRCKYMDVOPDKGAPYGGHLLNTLEKSRVHONLGENRPHITFOLLEGGEDILR  
 RLGLDNANOYOLIKGQCARVSSINRNDKRVRRALSTINNDIDIELLSIASV  
 LHLGNQAFADHGHAVYTNQIKYLRALNDKRVRRALSTINNDIDIELLSIASV  
 NLEQAVYADALAKAIYGRTPSWLVSKINSKYGKIDHKLKGSALVGLDITGFEEV  
 FQHSSEPOECINFCNEKLOQLFELTLCSEBODYVNHGHVTHGLDOKTKRVLGDER  
 KRKCTISLDECELRPGEDMTFLEKLEPTVANHGHVTHGLDOKTKRVLGDER  
 LHTAGVSNVAVGLFDRKNDLDFRNLKOPARDEVLIAHQVFNSEKLRKRETA  
 AGFAVRRKYEFLQRYKSLCPDPTPMWMDRADGVAFLVLSLGSVPEPKYAKRKIEI  
 RRPRTLFATEDALEVRKHSTATFLQARMGYHQHOKFLHMKHSVPEPKYAKRKIEI  
 KAARKMAVDVREFTKGATYRNQPCRTENEFLLDYIKYSLMLTRNOKSVLDKSM  
 VPPPSLRASSELREKCMNNMVKYCRINPEKMOOLEKVAASETFKOKNDYPSM  
 VPRFLINTRGNDENIKTILOLESQTLTAVAPVYKRYKRFKRRQLDLITONAYL  
 VEEAKMKORADIANLNGISVSLSDNLVHVKCEDNKQGDALLOSDDHYIELTLVVA  
 ITAEKININININGSLIKFVPGKEGCIIDTFASELILIAAKNGHLSVAVPRLSR"  
 2161. 2367  
 /note="3 tandem Ilegln motifs; encodes putative calmodulin  
 binding domains" 607 c 777 g 835 t

BASE COUNT 1009 a 607 c 777 g 835 t

ORIGIN

Query Match 47.1%; Score 1724.4; DB 5; Length 3288;  
 Best Local Similarity 71.8%; Pred. No. 0;  
 Matches 2270; Conservative 0; Mismatches 891; Indels 1; Gaps 1;

QY 12 CCGGGTCGGCAGAGATGCCCTACCGGGCATCGACCTGGGACAGTACGAGGGTTCAGTGA 71  
 Db 6 CCGGGTGGAGTGCGATAGTGGGACACCTTCAGGCTCCAGCAATCGAGGAAATCCAAATA 65  
 QY 72 CCATGAGAGCCCTTGTGACTGCGGAGACCGGGAGGGGTGACAGACTTGTCTGTCTG 131  
 Db 66 CCATGAGAGTGCCTTAACACCCGAGATGGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 125  
 QY 132 ACAAATTCACAGTACGAGTGCCTTCATTTAGAACCTCCGGCGGCGGTTCGGGAGAAC 191  
 Db 126 AAAACTACACAGTGAAGCTGCATTTATAGAAAACCTGAGAAAGGATTTAAAGAAATC 185  
 QY 192 TCATTTATACCTATATGCTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 251  
 Db 186 TCATATATATACGTAATTTGGGTGATGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 245  
 QY 252 TCTACAGCCCGCAGCATATATGAAACCTACCGGTGTCTGTCTGTCTGTCTGTCTGTCTGT 311  
 Db 246 TTTACGACAGACAGACATGAGAGATATGTTGAGATTTCTATGAAAGTCTCTCTC 305  
 QY 312 ATTTGTTTGCAGTGTGCTGACACTGTATACCGGGACCTGTCTGTCTGTCTGTCTGTCTGT 371  
 Db 306 ATATATATGCTATACAGATATACCTCTACCGCTCTTGGCACAGAGAGAAACACAGT 365  
 QY 372 CAGTATGATTTTTCGAGAGAGTGGGCGAGGCAAGACAGAGCCCAAGAGACTGTCTC 431  
 Db 366 GATATCTTATATATGCTGTGAGAGAGTGGGCTGGAGAACTGAGGACAGCAAGAAATCTC 425  
 QY 432 AGTTCTATGAGAGACTGTGCGCCCTGGAACGGGGTGGCGAGTGCAGACCGGCTGT 491  
 Db 426 AGTACTATGCTGTGACGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 485  
 QY 492 TCGAGACACCCCGTGTAGAGCGCTTGGGAATGCCAAGACTCTCCCAAGCAATACT 551  
 Db 486 TCGAGTCAATCTGTGTGGAGGATTTGGCAATGCAAAACCTTATGAAATGTAAATTT 545  
 QY 552 CCAGCCGGTTCGAAATGATCATGTGATGCTGACCTTGAAGGTGCGCCGAGAG 611  
 Db 546 CCAGCCGCTTGGAAATATATGATGTGCAATTTATTTAACAAGGACCTCTCGGTG 605  
 QY 612 GGCATATCTCAGTTACCTCTCTGAAAGAGCCCGGTGTGACCAAAATGACGAGAG 671  
 Db 606 GACATATTTTAACTATCTCTTGGAAAAGTCCCGTGTGATCAGAACACGAGAG 665  
 QY 672 GGAATCTTCAAGCTTTCACAGCTACTGAGAGGGGCGAGAGAGAGACTCTCTCGTGC 731  
 Db 666 GAAATCTTCAATTTTTCACAGCTTTCGAGGTGAGGAGGAGAAAGAGATTTGTAACGAG 725  
 QY 732 TGGGCTTGGAGAGGAAACCCCAAGCTACTTGTACCTGTGCAAGGSCAGTGTGCAAG 791  
 Db 726 TGGGCTGTGATTAATATGCCAAATTTACAGATCTGTGATTAAGGTGTAGTGTGCAAG 785  
 QY 792 TCTCTTCATCAACAGAGAGTGTGAGAGGTTATGAGAGAGGCGGTGCTGTGATTTG 851  
 Db 786 TTAGCTCATATGAATTAATAATGACGTGAGAGGTTGTTAGGACCTTGTCAATTTTAA 845  
 QY 852 ACTTACAGAGAGTGAAGTGAAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 911  
 Db 846 ACTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 905  
 QY 912 GCAACATCCACTTGTGCTGTGAGAGAGAGTATGCCAGTGTACTGTGAGAGAGAGAG 971  
 Db 906 GAAATGTGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 965  
 QY 972 TCAATATATGACAGAGGCTTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAG 1031  
 Db 966 TTAATATCTGACAGGCTCTGTCTGTGATGACTGTGATGAGGAGTGTCTGTCTGTCTGTCT 1025





```

RESULT 10
LOCUS      RC014549
DEFINITION Rana catesbeiana myosin I beta (AMI beta) mRNA
ACCESSION  U14549
VERSION     014549.1
KEYWORDS   GI:602137
SOURCE     Rana catesbeiana.
ORGANISM   Rana catesbeiana.
REFERENCE  1 (bases 1 to 3452)
AUTHORS   Metcalf, A.B., Chelliah, Y. and Hudspeeth, A.J.
TITLE     Metcalf, A.B., Chelliah, Y. and Hudspeeth, A.J.
          Molecular cloning of a myosin I beta isozyme that may mediate
          adaptation by hair cells of the bullfrog's internal ear
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 91 (25), 11821-11825 (1994)
MEDLINE    95083594
PUBMED     7991542
REFERENCE  2 (bases 1 to 3452)
AUTHORS   Metcalf, A.B.
TITLE     Metcalf, A.B.
          Direct Submission
JOURNAL    Submitted (08-SEP-1994) Anne B. Metcalf, Cell Biology and
          Neuroscience, University of Texas, 5323 Harry Hines Blvd, Dallas,
          TX 75235, USA
FEATURES
  source
    1..3452
    location/Qualifiers
      /organism="Rana catesbeiana"
      /db_xref="taxon:8400"
      /issue_type="brain"
      /gene="AMI beta"
      /gene="AMI beta"
      /gene="AMI beta"
      /gene="AMI beta"
      /note="partially processed transcript"
      /codon_start=1
      /product="myosin I beta"
      /protein_id="AA57192.1"
      /db_xref="GI:602138"
      /translation="MESALTRADRVQVDVLENTSEAFTELRKRRENLITTY
      IGSVLAVNPYKELEIYSKOHMERVRSVSEYSPHIVAIADNSYSLRPTKRDCL
      ISGSGACGTAEKSKILQYVAVTCVSDQVEYKADRLQSNHIVLAFKAGKILNDNS
      SRFQKYMVDQYRGAPRGAGHILNLLKESRVVHONGHGRNHITVQLLEGEDLAR
      RLGLDKNQNQOYLTKOCARVSSINDKNDKVRVRLASTINFNDDILELISVAV
      NLEQAVARDAALAIKGRFSWLVSKINKSLAYGTPRESLHKKIITAKGEELNSPL
      FOHNSFQPCINFCNEKIQLFELTKSEODEYSEGLAMEPVQYENNKITCDLVE
      LKHAGEVNSVAGFLDKNNDLFLRLKRVKCDSDSNPIAHOCFNRSELTDRKRDPEP
      TQFNLSLSKLMELIKOPSYVRCIPNDAKOPAFEDVILRHQVKTLGILNVRVRA
      RFPKTLFATEADALEVRKHSIATFLQARRGRHADQVAVLKLKSLGKPEYMGKTFEI
      KAARKWAVDVVRRFKGTIRNOPRCRNEYFLDYIRSEFLMIVLIRNOKSVLQKRS
      VPPSLREASELREKMNWKKYCRINREYFADYIRSEFLMIVLIRNOKSVLQKRS
      VPRFLFNTRLGDELINIKLQLESQTLTYAVPVKKDRGYKFREROLLQNMAYL
      VERAKKORIDANITGTSVSLSDNLFVLHVCENKOKGDITLQSDHVIETLRVA
      ITAEKKNININIGSIRKFTVGPKEGIIIDTFASSELLIAKNGHLISVAVPRLNR"
BASE COUNT 1056 a 695 c 821 g 880 t
ORIGIN

```

Query Match 44.6%; Score 1632.8; DB 5; Length 3452;  
 Best Local Similarity 70.0%; Pred. No. 0;  
 Matches 2269; Conservative 0; Mismatches 892; Indels 81; Gaps 2;

```

12 CCGGCTCCCGCAGATCGCGTACCGGCGATCGCCCTGGGCGATGACGGGGTTGAGTGA 71
Db 1111111111111111111111111111111111111111111111111111111
QY 52 CCGGCTGAGATCGCATAGTGGCAGCGCTTACGGCTCCACCAATCGACGGATCGAATAA 111
Db 1111111111111111111111111111111111111111111111111111111
QY 72 CCATGAGAGCGCGCTGATCCCGAGACCGGCTAGGGGTGGAGACTTTTCTCTGCTGG 131
Db 1111111111111111111111111111111111111111111111111111111
QY 112 CCATGAGAGATGCTTACAGAGCCCGGATCGGGTGGTGTCCAGAGACTTGTCTGCTGG 171
Db 1111111111111111111111111111111111111111111111111111111
QY 132 AGAATTCACCAAGTGAAGCTGCTTCATTTAGAACCTCCGGCGGCTCCGGAGAAC 191
Db 1111111111111111111111111111111111111111111111111111111
QY 172 AAAACTACACAGTGAAGCTGCTATTTAGAAAACCTAGAAACGATTTAAAGAAATTC 231
Db 1111111111111111111111111111111111111111111111111111111
QY 192 TCATTTATACCTACATCGGTCTGCTGCTAGTCTCTGTCAATCCCTACCGAGACTAGA 251
Db 1111111111111111111111111111111111111111111111111111111
QY 232 TCATATATATAGTACATTTGGGTGAGCTTGTGTATGTCAATCTTACAAAGAGTTAGA 291
Db 1111111111111111111111111111111111111111111111111111111
QY 252 TCACAGCCGCGAGCATATGAGAGCTACCCGTGTGCTCAATTTCTATGAATCACCTC 311
Db 1111111111111111111111111111111111111111111111111111111
QY 292 TTACAGCAAGCAACACATGAGAGATATGTGAGTATGTTTCTATGAATGTCCTCTC 351
Db 1111111111111111111111111111111111111111111111111111111
QY 312 ATTTGTTGCAATGCTGCTGACACATGTATACCGGGGCACTTGTACTAGAGCTGGAG 371
Db 1111111111111111111111111111111111111111111111111111111
QY 352 ATATATATGCGATGACAGATTAACCTACCGCTCTCTGGGCAAGAGAACCACTTC 411
Db 1111111111111111111111111111111111111111111111111111111
QY 372 CAGTATGATTTTCGAGAGAGTGGGGCAGGCAAGAGAGCCAGCAAGAGACTGCTCC 431
Db 1111111111111111111111111111111111111111111111111111111
QY 412 GTATTTTATATCTGTTGAGAGTGGGGCTGGGAGATGAGAGCCAGCAAGAAATCTAC 471
Db 1111111111111111111111111111111111111111111111111111111
QY 432 AGTTCTATGAGAGACTCTCCAGCCCTGAAACGGGGTGGCGGAGAGACCGCTGT 491
Db 1111111111111111111111111111111111111111111111111111111
QY 472 AGTACTATGCTGTGACGTGCTCTGACGTGACAGGTGGAGACTGTGAAAGCCGACTTC 531
Db 1111111111111111111111111111111111111111111111111111111
QY 492 TGCAGAGCAACCCGCTGTAGAGGCTTTGGGAAATGCGCAAGACTCTCCGACATTA 551
Db 1111111111111111111111111111111111111111111111111111111
QY 532 TGCAGTCCATCTGCTTTTGGAGCATTTTGGCATCAAAAACCTTACGAATGATTA 591
Db 1111111111111111111111111111111111111111111111111111111
QY 552 CCAGCCGGTTGGAAGAAGACATGATGTGCAATTTGACTTCAAGGGTCCCGCTGGAG 611
Db 1111111111111111111111111111111111111111111111111111111
QY 592 CCAGCCGCTTGGAAATATATGATGTGCAATTTGACTTCAAGGGTCCCGCTGGAG 651
Db 1111111111111111111111111111111111111111111111111111111
QY 612 GCCACATCTGAGTACCTCTTGAAAAGTCCCGGGTGGTCCACCAAAATACGGAGAG 671
Db 1111111111111111111111111111111111111111111111111111111
QY 652 GACATATTTAACTATCTCTTGGAAAAGTCCCGTGTGTACATCAAGAACCGAGAGA 711
Db 1111111111111111111111111111111111111111111111111111111
QY 672 GGAATCTCCAGTCTTTTACAGCTACTGAGGGGGCGAGAGAGACTCTCCGCTGG 731
Db 1111111111111111111111111111111111111111111111111111111
QY 712 GAACCTCCATATTTTTCACAGCTTCTTGGAGGTGGAGAGATTTGCTACGAGAG 771
Db 1111111111111111111111111111111111111111111111111111111
QY 732 TGGGCTTGAAGAGAAACCCGAGAGCTACTTGTACCGGTGAAAGGGCCAGTGTGCAAG 791
Db 1111111111111111111111111111111111111111111111111111111
QY 772 TGGGCTGATTAATAATCCCAAAATTAACCAATTAAGGTGAGTGTCCAGAGG 831
Db 1111111111111111111111111111111111111111111111111111111
QY 792 TCTCCATCAACAGCAAGAGTACTGGAAGTTATGAGAGAGCGCTGCTCATTTATA 891
Db 1111111111111111111111111111111111111111111111111111111
QY 832 TTAGCTCAATCAATGATTAAGAGATTTGAGAGAGCGCTGCTCATTTATA 891
Db 1111111111111111111111111111111111111111111111111111111
QY 852 ACTTACTGAGAGATGATGAGAGACTTGTGAGATGTTGAGAGAGCGCTGCTCATTTATA 891
Db 1111111111111111111111111111111111111111111111111111111
QY 892 ACTTCAATGATGATGATTAAGAGATTTGAGAGATTTGAGAGAGCGCTGCTCATTTATA 891
Db 1111111111111111111111111111111111111111111111111111111
QY 912 GCAACATCCCTTGTGCTGACAGAGAGCAACATGCGCAGGTACTACTGAGAACAGC 951
Db 1111111111111111111111111111111111111111111111111111111
QY 952 GAATGTGAGTTTGCACAGATGAGCATGAGCATGCGCAAGTCAACAGAAAGAACAG 971
Db 1111111111111111111111111111111111111111111111111111111
QY 972 TCAATATCTGACAGGCTCTGTGTTGAGAGTACAACTTATGAGAGAGCGCTGAGACC 1011
Db 1111111111111111111111111111111111111111111111111111111
QY 1012 TTAATACCTGCGCAGGCTCTCTGTAGACTCTACTTACTAGGAGAGCTCTCATTT 1071
Db 1111111111111111111111111111111111111111111111111111111
QY 1032 ACAGAGATGATGATGAGAGAGAGAGAGTCTGAGCCCACTAGCAATTAACCTGAGAG 1091
Db 1111111111111111111111111111111111111111111111111111111
QY 1072 ACAGAGATGATGATGAGAGAGAGAGAGTCTGAGCCCACTAGCAATTAACCTGAGAG 1091
Db 1111111111111111111111111111111111111111111111111111111
QY 1092 CATATGCAAGGATGCGCTTGCAGAGCTGTGTACAGCGGAGATTAACCTGAGAG 1131
Db 1111111111111111111111111111111111111111111111111111111
QY 1132 CTATGCTGAGAGTCTTGGCCAAAGCAATCTATGCGCGTACTCTCTGCTGTGTA 1151
Db 1111111111111111111111111111111111111111111111111111111
QY 1152 GAAGATCAATAGCTACTGCGCTTAAAGAGCGTGAAGACCCCACTGAGCAAGAGACA 1211
Db 1111111111111111111111111111111111111111111111111111111
QY 1192 GCAAGTAAACAATCTCTGGCATCAAGGTACAGATATGACAAAGCTTGGAGAGCGAT 1251
Db 1111111111111111111111111111111111111111111111111111111
QY 1212 CGGTCTTGGCTCTGAGCAATTAAGGCTTTGAGAGTGTACAGCAATACAGCTTTCGAGC 1271

```



Db 1232 CTGTCATAGGACTTTTGGACATATATGTTTGAAGTTTTCACATATACAGCTTCGAGAC 1311  
Qy 1272 AGTTCTGATCAACTACTACATGAGAGCTGCAGCAGCTCTCTTCATCAGCTACTCA 1331  
Db 1312 AATTTTGCATTATTTTGGCATATGAAAACTGCAGCAGTTGTTATACAGTGCAGTTGA 1371  
Qy 1332 AGTGGAGCAGGAGAAATACAGAGCTGAGGCGATGCGTGGGAACCTTCAGTACTTCA 1391  
Db 1372 AGTGCAGCAGGATGATATGATATGTAAGGATTTGCATAGGAGCAGTCCAGTATTTTGA 1431  
Qy 1392 ACAACAGATCATCTGTACCTGGTAGAGAGAGTTCAAGGGCATTCATCTCATCTTGG 1451  
Db 1432 ATACAAATATATCTGCATCTGTGCAAGAGAAATTCAAAGGCGATCATTTCTCATCTGG 1491  
Qy 1452 ATGAAGATGCTGCTGGGGAGGCGAGGAGCTGACCTTTCTGGAGAAGTTGGAGG 1511  
Db 1492 ATGAAGATGTTTACGTTCAGAGAGAGCCACAGATATGACATCTCTGGAGAAGTTGGAGG 1551  
Qy 1512 ACACGTGCAGCCCGCCACCTCCTCTGACGACACAGCTGCTGACCGAGAGAGACAGGA 1571  
Db 1552 ACACGTGTAAGAACCCACACCTTGTGACACACAACTTGGAGACCGAAGAACCCGAA 1611  
Qy 1572 AATCCTTAGACCGAGGAGGATTCGCGCTTCTGCAATTATGCTGAGAGGTGACCTACAGTG 1631  
Db 1612 AAGTCTTGGCAGAGATATATCTGCTTGTGCTGCAATTATGCTGAGAGGTGAACTACAGTG 1671  
Qy 1632 TGACTGGTTCCTGATTAACAAATCAATGACCTCTCTCCGGAACCTGAGAGGAGACATGT 1691  
Db 1672 TACGAGGCTTCCTGTATATAAACAATGATCTGCTTTGCGCATCTGAGAGGAGTAAATGT 1731  
Qy 1692 GCAGCTCAATGAACCCCATCATGAGCCAGTGTGTTGACAGAAGAGCTCACTAGTGAAGA 1751  
Db 1732 GCGACTGTGAAGAACCCGATGTGACATCACTGTTTAAATCGCTGACATGACGAGCAAGA 1791  
Qy 1752 AGCGGCCAG-GACGCTGGCCACCCAGTTCAAGATGAGCTCTGAGAGCTCGTGGAGTCC 1810  
Db 1792 AAGGCCAGAGCGGACACACAGTAAATAATAGCTTCAGAAAGCTGATGGAGAAATTC 1851  
Qy 1811 TGAGGCTAAGAGAGCTGCTATATCGGCTGATCAAGCAAGAGCGCAAGCGCGG 1870  
Db 1852 TTATGTCAGAGAGCTTCATATCGTAGGCTGATTAAGGCCATATAGCAACAAACG 1911  
Qy 1871 GTGCTTTGATGAGTGTCTCATCCGACATCAGTGAAGTACTCGGAGTATGAGAAATC 1930  
Db 1912 CTCGTTTGTATGAATGCTGTATCAGACACAGGTGAATATGTGGACTTATTTGAATG 1971  
Qy 1931 TGCGGCTGCGAGAGCTGCTTGTGCTATCTGCGCAATATGAGGCTTCTGCA- 1985  
Db 1972 TGCGGCTGCGGAGCTGATTTGCTATAGGCGAAATATGAGATCTTCTTCAAAAGT 2031  
Qy 1986 ----- 1985  
Db 2032 TGAGAACAGTGAATGTTGTGTATGAAAGGACATTCGCCCAATATATATGATATAT 2091  
Qy 1986 -----GAGTACAAAGTCACTGTGCCAGAGACATGAGCCCATGAGGAGG 2030  
Db 2092 AATATTTGTTTCTCTAGTACAAATCTCTTGGCCCTGATCTTGGCCCTAACTGGGATGG 2151  
Qy 2031 AGCGCCAGAGATGTTGGCGCTGTGGTGCAGACACCTCGGCTCAAGCCAGAGAGTA 2090  
Db 2152 CCGTGCATGATGCTGTGGCTGTTTGTGTCAGAGACCTGGGGGTATAAACCGAGAGTA 2211  
Qy 2091 CAAATTTGGGAGAGTAAATGATCTTCAATCCGAAATCCCAAGCTATTTGCCACAGAGA 2150  
Db 2212 CAAATTTGGGAGAGCAAAATATTTCAATTCGTTTCCCAAAACCTTTTGGCAACAGAGA 2271  
Qy 2151 CTCCCTGGAAGTCCGCGCAGAGTCTAGCCACCAAGATCCAGGGGCTGAGAGGGCTT 2210  
Db 2272 TGCTTTAGAGAGAGAGAGACACAGATAGCAATCTCTCCAGGCTAGTGTGAGAGAGTA 2331  
Qy 2211 TCATTGGCGACGAAATTTCTCCGGGTGAAGGATCAGGCATCTGTATCCACTCATGCTG 2270  
1111 1111111111 11 1111 11 11 111111 1111

Db 2332 TCATCAAGCAGAAATTCCTGACATGAACATTCGCCGGTAGAGATCCAGTCTGTG 2391  
Qy 2271 GCGTGGCACACTGGGCGGAGAGAGGACCAAGAGAAAGTGGGAGGCCACCAACCATCG 2330  
Db 2392 GCGTGAACAATAGCAGACGAGAAAGCTGCAAAAAGAAAGTGGCTGTGGATGTGTTGG 2451  
Qy 2331 TCGACTCATCCGTGCTTCATTTTGGCCATTCACCCCGGCTCCCTGAGATATGCTTCTT 2390  
Db 2452 GCGTTTATCAAAAGGCTTCATTTACCGCAATCAACCCGTTGCTACGTGAGAAATGACTT 2511  
Qy 2391 CTGGACCAACGTCGGGCTGCTATTTTGTATACCTGAGGAGGCGCAAGTCCCGCAAT 2450  
Db 2512 CTTGACATCATTTCTCTACTCTTCTTATGACCTCTTATGCGAATTCAGCAAAAGTGT 2571  
Qy 2451 TCTGACACCTCTCGGCCACACCCACCTGCTGAGAGAGGCTCAGAACTCTACG 2510  
Db 2572 GCTGACACAGTCTGTGGCTGTGTCACCTTCATCGCTCAGAGAGGCTCTGAGCTCTGAG 2631  
Qy 2511 GGAAGTGTGATGAAGAACATGTTGGAAGTACGCGGAGCATCAGCCCTGAGTGAA 2570  
Db 2632 GGAATGTGATGAACACATGTTGGAAGTATGCGCAGAAATTAACCTGAATGGA 2691  
Qy 2571 GCAGAGCTGCAGCAAAAGGCGGTGGCTAGTGAATTTTCAAGGCGCAAGAGACACTA 2630  
Db 2692 GCAACAGTTGGAACAGAAAGTTGTCAGTGAATATTCAGAGCAAAAGCAAAATTA 2751  
Qy 2631 CCCCAGAGTGTCCCGACACTCTTCAATGACACAGCGCTTGGCACAGAGAGATCAGCCC 2690  
Db 2752 TCCGAGAGTGTTCACAGGCTTTTATTAACACAGACTTGGAAATGATGAATCAACAC 2811  
Qy 2691 CAGAGTGTTCATCTTGGGCTCGAACCCATCCAGATGAGCGGTCGGTAAATA 2750  
Db 2812 TAAATTTCTTACGAAATTGGAAGCCAGACATTAACGTATGACAGTCCGTAGTGAAGTA 2871  
Qy 2751 CGACCGTAAGGTTTCAAGGCTCGCCCGGAGAGTGTGCTGACAGGCCAGGCTGTGCT 2810  
Db 2872 TGACAGAAAGGGGTACAAAGCTCCGGAAGAGACAACTTCTGCTACCTCAAAATGCTGCTA 2931  
Qy 2811 CATTGTGAGAGATGCTAAAGTCAAGCAGAGATTTGATTTGCAACCTAACCGGAATTC 2870  
Db 2932 TCTTTGTGAGGAAGCCAAATATAGCAGAGAGTTGACTATGCTAACCTACAGGAATTC 2991  
Qy 2871 TCTAGTACCTGAGTATGATAGCTATTTGTGCTTACGCTGACGCTGAGAGACAAAGA 2930  
Db 2992 CGTACAGAGTCTAAGTGAACACCTGTTGTCTCCATGTTAATGAGAGACAAACAA 3051  
Qy 2931 GAAGGAGATGTGTGCTGCAAGATGATGATCCAGACACTBACCAAGAGCGCTT 2990  
Db 3052 AAGGAGAGATGCAATTTGCAAGAGGATCATGTGATTAACACTGACAAAGTTGCCAT 3111  
Qy 2991 CAGTGTGACCGCGTGAACAATATCAACATCAACAGGCGACGATTAAGCTTGGCAGGGG 3050  
Db 3112 AACTCTGAGAAATATCAATATATCAACTATATAGGCGACACATCAACTTTACAGTTGG 3171  
Qy 3051 TCCAGGACAGGAGCATATGACTTCAATCGGAGCTTCAGAGCTTCTCATCACAGGCG 3110  
Db 3172 TCCAGGTAAGAGAGATATGACTTACTGCGGCTCTGAGTTACTATATAGCCAAAC 3231  
Qy 3111 TAAGAAATGCGCACTGTGCTGTGTGGCCCCACGCTGAATTTCTCGGTATGAAGGCTGG 3170  
Db 3232 AAAAAATGACACTTGTGTGTGTGCTCTCGTGTGAATTCACGATGATGACACTGCA 3291  
Qy 3171 GT 3172  
Db 3292 GT 3293

RESULT 11  
AY069044 3445 bp mRNA linear INV 17-DEC-2001  
LOCUS  
DEFINITION Drosophila melanogaster GH04201 full length cDNA.  
ACCESSION AY069044  
VERSION AY069044.1 GI:17861423

## KEYWORDS

FLI.CDNA.  
Drosophila melanogaster.

## SOURCE

Drosophila melanogaster.

## ORGANISM

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

## REFERENCE

1 (bases 1 to 3445)  
Stapleton, M., Brokstein, P., Hong, L., Abmayan, A., Carlson, J., Gonzalez, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Nuno, J., Pacleb, J., Paragas, V., Park, S., Phoumanavong, S., Wan, K., Yu, C., Lewis, S.E., Rubin, G.M. and Celisner, S.,  
Direct Submission  
Submitted (10-DEC-2001) Berkeley Drosophila Genome Project,  
Berkeley, CA 94720, USA

## AUTHORS

Lawrence Berkeley National Laboratory, One Cyclotron Road,  
Berkeley, CA 94720, USA  
Sequence submitted by:  
Berkeley Drosophila Genome Project  
Berkeley, CA 94720

## TITLE

This clone was sequenced as part of a high-throughput process to  
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,  
Science 2000). The sequence has been subjected to integrity checks  
for sequence accuracy, presence of a polyA tail and contiguity  
within 100 kb in the genome. Thus we believe the sequence to  
reflect accurately this particular cDNA clone. However, there are  
artifacts associated with the generation of cDNA clones that may  
have not been detected in our initial analyses such as internal  
priming, priming from contaminating genomic DNA, retained introns  
due to reverse transcription of unspliced precursor RNAs, and  
reverse transcriptase errors that result in single base changes.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our Web site  
(http://fruitfly.berkeley.edu) or send email to  
cdna@fruitfly.berkeley.edu.

## COMMENT

Location/Qualifiers  
1..3445  
/organism="Drosophila melanogaster."  
/strain="y; cn bw sp"  
/db\_xref="taxon:7227"  
/map="61F6-61F7"  
1..3445  
/gene="Myo61F"  
/note="alignment with genomic scaffold AE003471"  
/db\_xref="FLYBASE:FBgn0010246"  
117..3224  
/gene="Myo61F"  
/note="Longest ORF"  
/codon\_start=-1  
/product="GH04201P"  
/protein\_id="AL39189.1"  
/db\_xref="GI:17861424"  
/db\_xref="FLYBASE:FBgn0010246"

## FEATURES

gene  
CDS

## BASE COUNT

907 a 898 c 928 g 712 t

## ORIGIN

Query Match 25.1%; Score 919.6; DB 3; Length 3445;  
Best Local Similarity 58.6%; Pred. No. 1.e-220;  
Matches 1724; Conservative 0; Mismatches 1189; Indels 31; Gaps 6;

46 CTTGGGCACTGACGGGGTTCAGAGTACCATGAGAGGCGCTTACTGCCGAGACCGGCT 105  
116 CATGGCCAGCTTTTATTTCCCACTGAAATGAGAGCGGCGTCCAGACGCGATGCTGC 175  
106 AAGGGTGCAGACCTTTGCTGCTGAGAAATTCACCAAGGCGCTCTTCAATGAGAA 165  
176 TCGATCCAGACCTTTGCTGCTGAGAAATTCACCAAGGCGCTCTTCAATGAGAA 165  
166 CCGTCCGCGGCGCTTCCGAGAACTCTATTTACCTACATGCGTCTCTCCAGAGCA 235  
236 TCTGAGAGAGCGGTTCCAGAGAGCTGATCTACCTACCTATGCGGCTGTTGATCTC 225  
226 TGTCAATCCCTACCGAGCTTACAGATTTACAGCGGCGAGATATGAGACCTACCGTGG 285  
296 CGTGAATCCCTACAGACAGCTGCTCCATCTACCGAGATGACATGACGCGTACAGAAA 355  
286 TGTCACTTTCTTGAAGTACCACTATTTGTTGAGAGTGGCTGACATGTTACCGGGC 345  
356 CAAGCACTTTTGAAGTACCACTATTTGTTGAGAGTGGCTGACATGTTACCGGGC 345  
346 ACTTGTCTACAGGCTGCGGAGCGAGCTGATGATTTGAGAGAGTGGCGAGCAAA 405  
416 GCTGATCCAGAGAAACCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 415  
406 GACAGAGCGGCGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 475  
476 GACAGAGCGGCGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 465  
466 GGGTGGCGGCGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 535  
536 CCGTGGAGGCGGCGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 525  
526 TGGCAAGCTTCCGCAAGAGTACCTGAGCGGCTTGGAAATGATGATGATGATGATGATG 595  
596 TGGCAAGCTTCCGCAAGAGTACCTGAGCGGCTTGGAAATGATGATGATGATGATGATG 585  
586 TGGCAAGCTTCCGCAAGAGTACCTGAGCGGCTTGGAAATGATGATGATGATGATGATG 555  
656 CGATTTCAAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 645  
646 GGTGTCGACCAAAATACCGAGAGCGGAGCTTCCAGCTTTCAGAGTACCTGAGAGG 705  
716 AGTGGTGCCTCAATATGAGAGGAGCGGAGCTTCCAGCTTTCAGAGTACCTGAGAGG 715  
706 GGGGAGAGAGAGAGCTTCCAGCTTTCAGAGTACCTGAGAGGAGCTTTCAGAGTACCTG 775  
776 CGCGGAGAGAGAGCTTCCAGCTTTCAGAGTACCTGAGAGGAGCTTTCAGAGTACCTG 765  
766 CTTGGAGAGAGAGCTTCCAGCTTTCAGAGTACCTGAGAGGAGCTTTCAGAGTACCTG 835  
836 CCGAGAGAGAGAGCTTCCAGCTTTCAGAGTACCTGAGAGGAGCTTTCAGAGTACCTG 825  
826 TATGAGAGAGAGAGCTTCCAGCTTTCAGAGTACCTGAGAGGAGCTTTCAGAGTACCTG 895  
896 GGTGTCGACCAAAATACCGAGAGCGGAGCTTCCAGCTTTCAGAGTACCTGAGAGG 885  
886 CATGTCGACCAAAATACCGAGAGCGGAGCTTCCAGCTTTCAGAGTACCTGAGAGG 955  
956 AATGTCGACCAAAATACCGAGAGCGGAGCTTCCAGCTTTCAGAGTACCTGAGAGG 945  
946 TGGCAGAGTACTTGAAGACCACTCAAAATATCTACCAAGGCTCTTCTGAGTGAAGG 1012  
1013 TGGCAGAGTACTTGAAGACCACTCAAAATATCTACCAAGGCTCTTCTGAGTGAAGG 1005  
1006 TACACACTTGAAGAGCGGAGAGCTTGAAGAGAGTACCAAGGAGAGAGAGAGAGCTCT 1072  
1073 AAGCAAGTGAAGAGCGGAGAGCTTGAAGAGAGTACCAAGGAGAGAGAGAGAGAGCTCT 1065

1066 GAGCCACGTAACCTTGAACAGCGGCATATGCAAGGATGCGCTTGGCCAAAGCTGTGA 1125  
1133 CTCACACTGAACAGAGAGCTGGCATATGCCAGGAGCGCTTGGCCAAAGCTGTGA 1192  
1126 CAGCCGACATTCACCTGGCTGGTCAGAAAGATCAATAGCTACCTGCGCTTAAGAGCGC 1185  
1193 TGATGCTGCTCTGCTGGTGGTTCAGCGCTCAACATCTGCTGACAGGCGCAAGAAAC 1252  
1186 TGAGAGCCCGAGCTGGCAGACGCGTCTTGGGCTCCGCGACATTTAGGCGTTGA 1245  
1233 GCGGCATATC-----AGAAATATGTGATGGGCATTTGACATCTATGGCTTGA 1303  
1246 AGTGTTCGATTAACAGCTTGGAGAGTTCGATCACTCACTCAATGAGAAGCTGA 1305  
1304 GATCTTCAGAGAAACAGCTTTGAGCAGTTCGATCAATTTCTGCACAGACAGCTGA 1363  
1306 GCAGCTTTCATGAGCTGACTCTCAAGTCGAGCAGAGAAATACGAGCTGAGGCGAT 1365  
1384 GCAGCTGTCATGAGCTTACCTTAAGTCGAGCAGAGATGATACCGCGGAGGCGAT 1423  
1386 CGCGTGGGACCTGTCGATCTCAACAGCAAGATCATCTGACCTGATGAGAGAA 1425  
1424 CGAGTGGATACCTTGGAGTACTGACACAAAGGATTTGCAATCTGATGAGAGAA 1483  
1426 GTTCAAGGACATCATCTCCATCTTGGATGAAGAGTCCGCTCGGAGGAGCCACGA 1485  
1484 GCACAGGACCATCATCTCATCTGACGAGAGATGCTTGGCAGCAGAGAGCCACGA 1543  
1486 CCTGACCTTTCGAGAGAGTGGAGACACTGTCAAGCCCACTTCTGTCGACGA 1545  
1544 TAAGACCTTCTTGAAGAGCTCACCCAGAAATTTGGCCCGCACCATCTACGTTTGCA 1603  
1546 CAGGCTGCTGACCAAGAGACCAAGAAATCCCTAGACCGAGGAGGCTCGCTTCTGA 1605  
1604 CGAAAGGCGCCGCCACATCAAGAAATCATGCTGCGCATGAGTTCGCTGCTGA 1663  
1606 TTATCTGAGAGAGTGTGACTGTGACTGTGATGATGATTAAGAAATGAGCTCT 1665  
1664 CTATCGCGGAGAGTCACTTACAGTGTCAATGATTTCTGAGCAAGAAACAGACCTGTT 1723  
1666 CTTCGGAACCTGAAGAGAGACATGTGACGCTCAATGAACCCCATCATGCGGAGCTT 1725  
1724 GTTCAGGAGATCTTAAGAGAGACCTTGAAGGCTGGCAAGGCTATGTGAGAACTGTT 1783  
1726 TCACAAGAGTGAAGTCAAGAGACAGAGGCGGAG-GAGGTTGGCCACCCAGTTCAGAT 1784  
1784 CCGGAGAGAGAGCTGCGAGCTGAGAGCTCGGAGAGGCTATCACCCAGTTCGCGGC 1843  
1785 GAGCCTCTGACAGCTGAGAGATCTGAGATCTGAAGAGCTGCTATATTCGCTGAT 1844  
1844 ATCTCTTAACACATCTGATGAGATCTGATGTGCAAGAGAGCGAGGTATATTCGCTGAT 1903  
1845 CAGCCCAACAGAGCAG 1904  
1904 CAGGCAACAGAGCAG 1963  
1905 GAAATGACCTGAG 1964  
1964 CAAATGACCTGAG 2023  
1965 CAAATGACCTGAG 2018  
2024 AAGATATGAGCTTCTGAG 2083  
2019 CATGTGGAG 2078  
2084 CAAAGGCTCCCGGTGGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2143  
2079 GCGAGAGAGTCAAAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2138  
2144 CGAGGAAAGTACAGAGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2203

2139 TGCACAGAGAGATCTCCTTGAAGTCCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 2198  
2204 CGATACGAGAGAGATGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2263  
2199 CTGAGAGGCTTTCATTTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2258  
2264 CTGAGAGGAGATGATGAG 2323  
2259 CCAGTCATGATGAG 2318  
2324 GCGAGAGATTTGTCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2383  
2319 CCAGAGATCTGTCGATCATCTGCTGCTTCAATTTTGCATTCACCCGAGTGCCTGA 2378  
2384 CGATTAAGATTCGCTGCTTCAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2443  
2379 GAATGCTTCTTCTGAG 2438  
2444 CATGAGAGAGATTTATTTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2503  
2439 GCGCGGAGATGTCGAG 2498  
2504 GCGCAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2563  
2499 AGAAGTCTACGAG 2558  
2564 GCGCATCTTTCACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2623  
2559 CCCTGATGAG 2618  
2624 GCGCAG 2683  
2619 GAAAGCACTACCCCGAG 2678  
2684 GAGAGACATACGAG 2743  
2679 GAG 2729  
2744 GCACATTCAG 2803  
2730 TGGCGTCCCGTGGTAAATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2789  
2804 CAGTCTCTTCTGACAG 2863  
2790 GCTACAG 2849  
2864 GCTGAG 2923  
2850 TGC--CAACCTAACCGAGATCTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2906  
2924 GCGGTAG 2983  
2907 CGTGAG 2966  
2984 CATACCGCTGAG 3043  
2967 CGAG 2970  
3044 CGAG 3047

RESULT 12  
DMU07596 3882 bp mRNA linear INV 08-JUL-1994  
LOCUS  
DEFINITION Drosophila melanogaster myosin-IB mRNA, complete cds.  
ACCESSION U07596  
VERSION U07596.1 GI:466257  
KEYWORDS  
SOURCE  
ORGANISM  
fruit fly.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## SOURCE

## FEATURES

## SOURCE

## FEATURES

## SOURCE

## FEATURES

## SOURCE

## FEATURES

## SOURCE

## FEATURES

## SOURCE

## FEATURES

## SOURCE

## FEATURES

## SOURCE

## FEATURES

## SOURCE

## FEATURES

## SOURCE

## FEATURES

## SOURCE

## FEATURES

## SOURCE

## FEATURES

## SOURCE

## FEATURES

## SOURCE

## FEATURES

## SOURCE

## FEATURES

## SOURCE

## FEATURES

## SOURCE

## FEATURES

## SOURCE

## FEATURES

## SOURCE

## FEATURES

## SOURCE

## FEATURES

## SOURCE

## FEATURES

## SOURCE

## FEATURES

## SOURCE

## FEATURES

```

OY      2458   ACCTCCTGGCCACACCCCCCACTGCGCCTTGAGAGAGGCCTCAGAATCTCTCAGAAATCTT
Db      2467   AAGAGCTGGCCCCCATGTGCCCGGGGACACTGCGAGAGAGGCTCTCCGCATCTCTCACCGCTTG
OY      2518   TGCATGAAGAACAATGTGTGTGAAGTACTGCCGGAGCATCACCCTGAGTGAAGACGACGAG
Db      2527   CACCGCTTTCTACTTGCGCCAGATCTACCGCCCTTTAAACTGACGCCGACACAGAAAGAGCGCAA
OY      2578   CTGCAGCAAAAAGGCGGTGGCTAGTGAATAATTTCACAGGGCAGAAAGAACACTTACCCCAGC
Db      2587   TTCCGAGCTGAAGAGTCTGCGGCGAGAAGGTCTTTTAAAGGCGCAAGAAACAACTACGCGAGCG
OY      2638   AGTGCCCCCAGACACTCTTCATTATAGCACACAGCGGCTTGCGACAGAGAGAGA-----TCAGC
Db      2647   AGTGTGTCCACTTGGTTCTCAGAGAGAGACCGCATCCCCAAGGAACACATPCACAGCATCAAC
OY      2689   CCCAGAGTGCCTTCAATCTCTTGGGCTCTGAACCCATCCAGTATGCCGTGCCCGGTGTAATA
Db      2707   GACTTCGTGGCCAGACACCTTCCTGGCGAGGAGGAGAGCTTAGTACCATGCTCTCTGCACCAAG
OY      2749   TACGACCTTAAGGTTTACAAAGCCTTCGCCCCCGGCGACGTGCTGTCCAGCCCACTGCTCTG
Db      2767   TTCGATCGCGACCGGCTTAAAGTCCGCGACCGCTTTATCTGCTGAGCAACAAAGGCTATTC
OY      2809   GTCATTTGTGAGAGATGCTTAAAGTCAAGAGAGAGAATTTGATTTATGC---CAACCTAAGCGGA
Db      2827   TATGTCTCTCGACGCAACAACTACAGAGAGAGACACCGCCTGCGCTGACAAAGATGTCAC
OY      2866   ATCTCTGTCACTAGCCTGAGTGAATAGCCTATTATTGTGCTTCACGTGCACGCTGAAGACAAAC
Db      2887   TTCACGTTGACGAACCAACAAACGACGACCTGATGTGTCATCCCATACCCCTGCACCTGAAAA
OY      2926   AAGCAGAGGAGAGATGTGTGCTGTGACAGTGTATCATCTGTGATCGAC
Db      2947   AAGGACAAGGCGGACTGTATTCTGTATCTTCGCGGCATATATCGAG

```

RESULT 13  
AC012834/c  
LOCUS DEFINITION AC012834 melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, In ordered pieces.  
ACCESSION AC012834  
VERSION AC012834.1 GI:6223487  
KEYWORDS HTG; HTGS\_PHASE2.  
SOURCE Drosophila melanogaster.  
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyrroidae; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 116374)  
AUTHORS Adams,M. and Venter,J.C.  
TITLE Direct Submission  
JOURNAL Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA  
COMMENT This sequence was identified as CDM.10210646 by the submitter. For further information on this sequence you may e-mail to fl@celera.com.  
FEATURES  
    source                 location= "Drosophila melanogaster"  
                          /db\_xref="taxon:7227"  
BASE COUNT             31293 a 26648 c 26933 g 31500 t  
ORIGIN  
Query Match             20.5%; Score 749.4; DB 2; Length 116374;  
Best Local Similarity   57.1%; Pred. NO. 1.3e-177;  
Matches 1561; Conservative 0; Mismatches 1081; Indels 93; Gaps 7;

[illegible]

Db 34649 GATATGCTGCTTGGACTGGCGAAGAGACCTGCCACCAAGGTACTGATTAAGACGTGGC 34590  
 QY 2468 CCACACCCCCACCTGCCCTGAGAGAGCCCTCAGAACTGCTACGGAACTGTGCATGAAGA 2527  
 Db 34589 CCGATGCCCTGGGAGACCTCCGAGAGAGGCTCCGGGATCTTACCCGCCCTGCACCGCTCC 34530  
 QY 2528 ACATGGTGTGAAGTACGTCGGAGCATTCAGCCCTGAGTGAAGACGACGTCGACAAA 2587  
 Db 34529 ACTTGCCGAGATCTACGCCCTTTAAAGTACGCCGAGAGAGAGCAATTCGAGCTGA 34470  
 QY 2588 AGCGCGTGGCTAGTGAATTTTCAAGGAGCAAGACACTACCCCGAGATGTCCCA 2647  
 Db 34469 AGTCTCGCGGAGGAAGGCTTTAAAGGCAAGAAAGAACAGCAGCAGCTGTGTCA 34410  
 QY 2648 GACTCTTCAATAGCAACAGCGCTTGGCAGAGAGAGA-----TCACCCCGAGAGTGC 2698  
 Db 34409 CTTGGTCCAGAGAGAGCCGATCCCAAGAGACATCCAGCAGTCAACGACTTCGTGG 34350  
 QY 2699 TTCAATCTTGGGCTCTGAACCCATTCAGTATCCGTCGCCGTGTAATAATACGACCTA 2758  
 Db 34349 CCAGACCTTCGGCAGCAGCAGCTTAAGTACAGTCTTTCGACCAAGTTGATCGGC 34290  
 QY 2759 AGGTATCAAGACCTGGCCCCGCGCAGCTCTGCTCAAGCCGAGTGTGTGATCTGTGG 2818  
 Db 34289 AGCGCTACAAAGTCCGCGCAGCCTTTATCTCTGTGAGCAAGGCTATATATGTCCTG 34230  
 QY 2819 AGCATGCTAAAGTCAAGCAGCAAGATGATTATGCG---CAACCTAACCCGAGATCTGTCA 2875  
 Db 34229 AGGCAAGACCTCAAGCAAGCAAGCAGCCGCTGCGCTACAGCAAGATCGACTTCACCTGA 34170  
 QY 2876 GTAGCCTGAGTATAGCTTATTTGCTTTCACGTGACGGCTGAAGACAACAGCAGACAG 2935  
 Db 34169 CCAACCAACAGCAGCAGCTGATGTCATCCGATACCCGCTGAAAGAGCAAG 34110  
 QY 2936 GAGATGTGCTGCTGCAAGTATCATGATGATGAG 2970  
 Db 34109 GCGACTGATTTGATGATCATTCGCGCATATATGAG 34075

## RESULT 14

AC010007

LOCUS AC010007 175481 bp DNA linear INV 01-JUN-2002  
 DEFINITION Drosophila melanogaster 3L BAC RP98-17K17 (Roswell Park Cancer  
 Institute Drosophila BAC Library) complete sequence.

AC010007

AC010007.4 GI:21306589

KEYWORDS

SOURCE

ORGANISM

Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydriidae; Drosophilidae; Drosophila.  
 1 (bases 1 to 175481)

REFERENCE

AUTHORS

Muzny,D., Scherer,S., Adams,M.D., Holt,R.A., Evans,C.A.,  
 Gocayne,J.D., Taber,P., Williamson,A., Homsif,F.H.,  
 Dugan-Hochs,S.D., Sodergren,E.S., Hodgson,A.H., Chen,R.C.,  
 Ayele,M., Scott,G.S., Worley,K.M., Amaratunga,P.G., Brandon,R.C.,  
 Rogers,Y., An,H., Baldwin,D., Beeson,K.Y., Brown,M., Buhay,C.,  
 Busam,D.A., Center,A., Chen,G., Chen,Z., Clerc-Blankenburg,K.,  
 Davidson,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.E.,  
 Draper,H., Emery-Cohen,A., Ferrera,S., Gary,N.D.S., Houck,J.,  
 Hostin,D., Howland,T.J., Hume,J., Ibegwam,C., Jalali,M., Kovar,C.,  
 Liu,W., Mettel,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B.,  
 Nelson,K.A., Ndassa,Y., Nguyen,N., Perez,L., Pittman,G.S., Puri,Y.,  
 Scheeler,F., Shen,H., Strong,R., Tector,C., Wang,Q., Williams,S.M.,  
 Xiang,L., Zaveri,J.S., Zhou,J., Zorilla,S., Smith,H.O.,  
 Wheeler,D., Weinstock,G., Gibbs,R. and Venter,J.C.

TITLE

JOURNAL

REFERENCE

AUTHORS

2 (bases 1 to 175481)  
 Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J.,  
 Benton,M., Bimane,K., Blankenburg,K., Bonnin,D., Bouck,J.,

## TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE  
 1 (bases 1 to 175481)  
 Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J.,  
 Benton,M., Bimane,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
 Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
 Burck,P., Burckel,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,  
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
 Deem,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
 Dugan-Hochs,S., Durbin,K.J., Barnhart,C., Edgar,D., Edwards,C.C.,  
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
 Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,  
 Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,  
 Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,  
 Hollins,B., Homsif,F., Howard,S., Huber,J., Huik,S., Hume,J.,  
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,  
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,  
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,  
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
 Louissege,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
 Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,  
 Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,  
 Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,  
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,  
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenwo,S.,  
 Ogun,M., Okwunonu,G., Otagunye,N., Oviedo,R., Pace,A., Payton,B.,  
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,  
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojupokan,I., Rolle,M.,  
 Ruiz,S., Savery,S., Scherer,S., Scott,G., Shen,H., Shoostari,N.,  
 Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,  
 Stone,H., Sutton,A., Swatek,A., Taber,P., Tamerisa,A., Tamerisa,K.,  
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,  
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,  
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,  
 Watlington,S., Williams,G., Williamson,A., Wlezyk,R., Wooden,S.,  
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,  
 Weinstock,G. and Gibbs,R.

## TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE  
 2 (bases 1 to 175481)  
 Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J.,  
 Benton,M., Bimane,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
 Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
 Burck,P., Burckel,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,  
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
 Deem,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
 Dugan-Hochs,S., Durbin,K.J., Barnhart,C., Edgar,D., Edwards,C.C.,  
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
 Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,  
 Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,  
 Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,  
 Hollins,B., Homsif,F., Howard,S., Huber,J., Huik,S., Hume,J.,  
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,  
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,  
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,  
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
 Louissege,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
 Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,  
 Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,  
 Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,  
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,  
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenwo,S.,  
 Ogun,M., Okwunonu,G., Otagunye,N., Oviedo,R., Pace,A., Payton,B.,



Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L., L.,  
Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rotubackan, I., Rolf, M.,  
Rulz, S., Sawyer, G., Scherer, S., Scott, G., Shen, H., Shooshart, N.,  
Stinson, I., Sodergren, E., Sonak, T., Sparks, A., Stanley, H.,  
Tang, H., Sutton, A., Swalek, A., Taylor, C., Taylor, P., Tamerisa, A.,  
Thomas, S., Tansley, K., Vasquez, L., Vera, V., Villalón, D.,  
Walli, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
Washington, S., Williams, G., Williamson, A., Wlezyk, R.,  
Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S.,  
Zorrilla, S., Nelson, D.,  
Direct Submission  
Submitted (29-DEC-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 175481)  
BCM-HGSC.  
Direct Submission  
Submitted (01-JUN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jun 1, 2002 this sequence version replaced g1:17998562.  
COMMENT INFORMATION: [http://www.hgsc.bcm.tmc.edu/or\\_email](http://www.hgsc.bcm.tmc.edu/or_email)  
[gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)







```

OY 1515 CTGTCAAGCCCAACCTCTGACGACACAGCTCGCTGACCAAGAACAGAAAT 1574
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 117413 AATTGGCCAGACGACCACTGTTGGCAAAAGGCCGCCCAACATCAAGAAA 117354
OY 1575 CCTTAGACGAGGAGGATTCGCTTTCATTTATCTGAGAGGAGCTACAGTGA 1634
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 117353 TCATGCTGCGGATAGTTCCTGCTTGTCACCTATGCGGCGAGGTCACTACATGTTCA 117294
OY 1635 CTGGGTTTCTGGATAAAAACATGACCTCTCTCCGCAACCTGAAAGAGACATGTGA 1694
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 117293 ATGATTCCTGGACAAGAACAGACAGCTGTTGTTCAGGATCTAAAGAGACCTCAGGA 117234
OY 1655 GCTCATGACCCCAATCATGCCCCAGTGGCTTTGACAGAGTAGTCACTGACGAAGAAC 1754
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 117233 AGGCTGGAGACGCGCATTTGAGAGAACTGCTTCCGAGAAAGAGCTGGCCACACCTGAAGC 117174
OY 1755 GGGCAG-GACGGTGGCCACCCGTTCAAGATAGGCTCTCGAGCTCGTGGAGATCTGCA 1813
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 117173 GTCCGAGACGGCTATCACACCGTTCCGGGCATCTTTAACACCTCATGGACATTTCTGA 117114
OY 1814 GGTCTAAGGAGCGCTGCTATATCCGGTCATCAGCAACAAAGCGCCAAAGCGGGTTC 1873
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 117113 TGTGAAGAGCGCGAGCTACATATCCGCTGATCAGCAACAGACCTGCGAGCTGCCAACG 117054
OY 1874 GCTTTGATGAGGTGCTCATCCGACATCAGGTGAAGTACCTGGAGCTGATGAGAAATCTGC 1933
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 117053 TCTTCAACGATGATTTGGTACTACACAGTCAAGTACCTTGCGCTGATGAGAAATCTGC 116994
OY 1934 GGTTCGCGAGACCTGCTTTCGCTATCCGTCGAATATGAGCTTTCTGCGAGAGTACA 1993
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116993 GTGTAGGCGAGCTGCTGCTTCCACCGACGAACTATGAGCTCTTCCGAGCGCTTACA 116934
OY 1994 AGTCACTGTGCCAGACAGATGGCC-----CATGTGGCGAGACGGCCCGAGATGTG 2047
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116933 AGTCCCTGAGCAATGCCACCTGGCCCAATTACAAAGGTCCCGTGGCCGAAAGCGGTG 116874
OY 2048 TGGCCGTGTGGTCAACACCTGGCTACAGCCAGAAAGATACAAATGGCGAGACTA 2107
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116873 TCCAGCAGCTGGTGAAGATTTGGGCTGGGACGAGAAAGTACAGGTGGCGAGACGA 116814
OY 2108 AGATCTTCATCCGATTTCCCAAGACCTTATTTGCCACAGAGACTCCCTGSAATCCGGC 2167
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116813 AACTTATTCCTCGCTGGCGGAGACCTTGTGATACGAGAGATGCTACACAGAGAAAGA 116754
OY 2168 GGCAGAGTCTACCCACCAAGATCCAGCGGCTTGGAGGGGCTTTCATTGGCGACAGAAAT 2227
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116753 AACATGAGATACGGGCGCATTCACAGGCCCACTGGAGAGGATTTGCAAGAGAGAAAT 116694
OY 2228 TTCTCCGGGTGAAGCGATCAGCCATCTGTATCCAGTCACTGTGGCGTGGCACACTGGGCC 2287
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116693 ATCTGAAACTGCGTCTCAGTGTATCATATGACAGCTATTGTGCGCAAGTTGGGCAC 116634
OY 2288 GGAGGAAGGCGACCAAGAGAGTGGGCAAGCCAGACCATCCGTGACTCATCCGTGGCT 2347
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116633 AGCAGCGCGCCAAAGAGGCGAGGCGCGGATTAAGATTCGCTTATCAAGGGCT 116574
OY 2348 TCATTTTGGCCATTTCACCCCGGCTGAGATAGCCCTTCTTGGACACAGTGGCGG 2407
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116573 TCATACCCCGAAGCATGCCCCGAATGGCTTCAATGAAGATTTATGCAACGCCCAAGC 116514
OY 2408 CCTCATTTTGGCTTAACCTGAGAGCGGCAACTGCCCGGAATGTTGACACCTCTCTGC 2467
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116513 GTATGTGCTGCTTGCAGTGGGAGAGGCTCCGCAATCTTCAACGCGCTCTCC 116454
OY 2468 CCACACCCCCACCTGCTCGAGAGAGGCTTCAGAACTGCTACGGAGACTGTGCATGAGA 2527
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116453 CCCATGCCCTGGGCACTGGCGAGAGGCTCCGCAATCTTCAACGCGCTCTCC 116394
OY 2528 ACATGGTGTGAAGTACTGCGGAGCATTCAGCTGAGTGAAGCAGCAGTGCAGANA 2587
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116393 ACTTGGCCAGATCTTACCGCCTTAAACTGACGCCGAGAGAGAGCAATTCGAGCTGA 116334
OY 2588 AGCGGTGGCTAGTAATTTTCAAGGGCAAGAGACAACTACCCCAAGAGTGTCCCA 2647

```

```

    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116333 AGTCTTGCGGAGAGAGGCTTTAAGGCGCAAGAAACATACCGAGCACTGTGTCCA 116274
OY 2648 GACTCTTATGACACAGCGCTTGGCACAGAGAGA-----TCAGCCCCAGAGTGC 2698
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116273 CTTGGTTCCAGAGAGACCGCATCCCCAAAGAGACATCAGGAGTCAACGACTTCGTGG 116214
OY 2699 TTCAATCCTTGGGCTCTGAACCCATCCAGTATCCGTGCGCGTGGTAAATATGACCGTA 2758
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116213 CCAGACCTTGCGCAGCGAGCAGCTTAAGTACAGTCTTGTGCAACAAAGTTGATCGGC 116154
OY 2759 AGGTTACAGCTTGGCCCCCGGCGAGCTGCTCTACGCCCAAGTCTGTGCTATTTGG 2818
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116153 ACGGCTAACAGTCCCGCGACCGCTTTATCTCTGTGAGCAACAGGCTATATGTCTCG 116094
OY 2819 AGGATCTTAAGTCAAGCAGAGATTAATATGCTG---CAACCTAACCGGAATCTCTGCA 2875
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116093 ACGGCAAGACTTACAAGAGAGACACCGCTCGCTAGACAAATCGACTTCACGCTGA 116034
OY 2876 GTAGCCTGAGTATAGCCATTTTGTGCTTACAGTGCAGCGTGAAGACAACAGAGAGG 2935
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116033 CGAACCACAGACAGACACTGATGTGATCCGATACCGCTGACCTGAAGAAAGACAGG 115974
OY 2936 GAGATGTGCTGCAAGAGTATCATGTGATGAG 2970
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 115973 GCGACTTGATCTGATCATTCGCGCATATATGAG 115939

```

Search completed: December 9, 2002, 14:35:45  
 Job time : 7405 secs.



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 9, 2002, 11:25:50 ; Search time 485 Seconds  
(without alignments)  
16999.105 Million cell updates/sec

Title: US-09-893-371-3  
Perfect score: 3661  
Sequence: 1 ggaagcgggcgccgcggtccg.....ccacctgcacclttagcca 3661

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues  
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N.Geneseq\_101002:\*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*

2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*

3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*

4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*

5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*

6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*

7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*

8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*

9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*

10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*

11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*

12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*

13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*

14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*

15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*

16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*

17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*

18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*

19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*

20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*

21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*

22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*

23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*

24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2759.4	75.4	4202	22	AA159147	Human polynucleoti
2	2549	69.6	3392	23	AA574358	DNA encoding novel
3	1492	40.8	3771	23	AA574359	DNA encoding novel
4	1144.8	31.3	1727	22	AA160933	Human polynucleoti
5	875.2	23.9	3406	23	ABL19515	Drosophila melanog
6	873.8	23.9	3449	23	ABL19515	Drosophila melanog
7	749.4	20.5	6483	23	ABL19550	Drosophila melanog
8	749.4	20.5	7057	23	ABL04914	Drosophila melanog
9	745.2	20.4	2560	22	AAH02967	Human shear stress

10	510.6	13.9	3874	23	ABL28593	Drosophila melanog
11	507.4	13.9	3400	22	AA540995	CDNA encoding nove
12	506.2	13.8	3332	22	AA540998	CDNA encoding nove
13	460.8	12.6	3617	24	ABK52205	DNA encoding human
14	450.6	12.3	2263	22	AA532774	Human secreted pro
15	385.2	10.5	4760	24	AA594850	Human DNA sequence
16	361.4	9.9	3588	23	AA584797	DNA encoding novel
17	342.4	9.4	7465	24	ABN96860	Gene #3358 used to
18	325	8.9	1605	21	AA576860	Human ORFX ORF2415
19	305.2	8.3	1636	22	AAH34495	Human colon cancer
20	305.2	8.3	1712	21	AA516208	Human prostate can
21	293.6	8.0	2092	17	AA535762	Human non-conventi
22	283.8	7.8	7516	22	AA51987	Human polynucleoti
23	283.8	7.8	8486	22	AA52971	Human polynucleoti
24	282	7.7	6733	23	ABL07105	Drosophila melanog
25	278	7.6	4459	22	AA552256	Human polynucleoti
26	276.4	7.5	4515	22	AA552272	Human polynucleoti
27	274.6	7.5	3757	24	ABN59671	Novel human coding
28	274	7.5	2026	23	AA580792	DNA encoding novel
29	260	7.1	1405	17	AA535761	Human non-conventi
30	253.6	6.9	2349	23	AA567617	DNA encoding novel
31	253.6	6.9	2711	24	ABK84726	Human cDNA differe
32	250	6.8	5094	23	AA578835	DNA encoding novel
33	249.6	6.8	511	24	ABN73548	Bovine embryonic g
34	246.4	6.7	7736	23	AA565910	DNA encoding novel
35	238	6.5	7794	24	ABN59632	Novel human coding
36	236.8	6.5	6175	17	AA534291	coding sequence fo
37	235	6.4	6607	22	AAH98321	Rabbit EST-derived
38	233.4	6.4	789	21	AA577011	Human ORFX ORF2566
39	227.6	6.2	2486	23	ABV22140	Human prostate exp
40	227.6	6.2	2486	23	ABV27980	Human prostate exp
41	227.2	6.2	6655	22	AAH99691	Human protein enco
42	226.6	6.2	6390	23	ABL12023	Drosophila melanog
43	224.2	6.1	5574	24	ABK35570	Gene MYH11 differe
44	219	6.0	6237	23	AA576357	DNA encoding novel
45	218.4	6.0	3779	23	AA571357	DNA encoding novel

ALIGNMENTS

RESULT 1	
AA159147	
ID	AA159147 standard; cDNA; 4202 BP.
AC	AA159147;
XX	
DT	22-OCT-2001 (first entry)
XX	
DE	Human polynucleotide seq ID NO 1350.
XX	
KW	Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW	leukaemia; ss.
OS	Homo sapiens.
XX	
PN	WO200153312-A1.
XX	
PD	26-JUL-2001.
XX	
PF	26-DEC-2000; 2000WO-US34263.
XX	
PR	21-JAN-2000; 2000US-0488725.
PR	25-APR-2000; 2000US-0552317.
PR	09-JUL-2000; 2000US-0598042.
PR	19-JUL-2000; 2000US-0620312.
PR	03-AUG-2000; 2000US-0653450.
PR	14-SEP-2000; 2000US-0662191.
PR	19-OCT-2000; 2000US-0693036.

```

PR 29-NOV-2000: 2000US-0727344.
XX
XX (HISE-) HYSEQ INC.
XX
XX Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D:
XX PI Wang J, Meng Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J:
XX PI Zhao Q, Zhou P, Goodrich R, Dermanac RT:
XX WPI: 2001-442253/47.
XX P-PSDB: AAM3991.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Claim 1: SEQ ID NO 1350, 10078pp: English.
XX
XX The invention relates to human nucleic acids (A157798-A161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cyostatic activity. The polynucleotides are
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localized neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilization of the activities such as: immune system suppression,
XX and thrombolytic activity, chemotactic/chemokinetic activity, haemostatic
XX assays for receptor activity, cancer diagnosis and therapy, drug screening,
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
XX Sequence 4202 BP: 919 A; 1228 C; 1267 G; 788 T; 0 other:
XX
XX Query Match
XX Best Local Similarity 75.4%; Score 2759.4; DB 22; Length 4202;
XX Matches 3116; Conservative 0; Mismatches 481; Indels 8; Gaps 6:
XX
XX QY 44 GCCCTGGCAGTGCAGGGGTTGAGTGCAGTGCAGGAGCGCCCTGACGTCGCCGAGACCG 103
XX DB 292 GCCCTGGCAGTGCAGGGGTTGAGTGCAGTGCAGGAGCGCCCTGACGTCGCCGAGACCG 103
XX QY 104 GTAGGGGTGCAGGACTTGTCTGCTGTGAGAAATTTCACAGTGAAGGCTGCTTCATTGAG 163
XX DB 352 GTGGGGGTGCAGGAAATTTCGTCGTGTGAGAACTTCACAGCGAGCGCCCTTCATTGAG 163
XX QY 164 AACCTCGCGGGGGGTTCCGGGAGAACTTCATTAACTCACTGCTGCTGCTGCTGCTGAG 411
XX DB 412 AACCTCGCGGGGATTTTCGGGAGAACTTCATTAACTCACTGCTGCTGCTGCTGCTGAG 411
XX QY 224 TCTGTCAATCCCTACCGAGACCTACAGATCTACAGCCGCGACATATGAGACGTTACCGT 283
XX DB 472 TCTGTCAATCCCTACCGAGACCTACAGATCTACAGCCGCGACATATGAGACGTTACCGT 283
XX QY 284 GGTGTCAATCCCTACCGAGACCTACAGATCTACAGCCGCGACATATGAGACGTTACCGT 283
XX DB 532 GGTGTCAATCCCTACCGAGACCTACAGATCTACAGCCGCGACATATGAGACGTTACCGT 283
XX QY 344 GCACTTGTACTGAGCGTGGGAGCAAGCAGTGAATTTCTGAGAGTGGGAGCGGACCG 343
XX DB 552 GCACTTGTACTGAGCGTGGGAGCAAGCAGTGAATTTCTGAGAGTGGGAGCGGACCG 343
XX QY 404 AAGACAGAGGCGCAACAGAGCTGCTCAGTTCATATGCAAGAGACCTGCGCCAGCCCA 463
XX DB 652 AAGACAGAGGCGCAACAGAGCTGCTCAGTTCATATGCAAGAGACCTGCGCCAGCCCA 463
XX QY 464 CGGGGTGGCGCAGTTCGAGACCGCGTGTGCAAGACCAACCGCGTGAAGGCTTTGGG 711
XX DB 712 CGGGGTGGCGCAGTTCGAGACCGCGTGTGCAAGACCAACCGCGTGAAGGCTTTGGG 711
XX QY 524 AATCCCAAGACTCCGCAACGATTAATCCAGCGGTTTGAAGATGATGATGAGGAG 583
XX DB 583 AATCCCAAGACTCCGCAACGATTAATCCAGCGGTTTGAAGATGATGATGAGGAG 583

```

---

```

DB 772 AATGCCAAGACCTTCGGAACGATTAATCTCAGAGCTTCGGGAAGTACATGATGTGCAG 831
QY 584 TTGTGACTTCAAGAGGTGCCCCCGTGGAGGCGCAATCTCACTTACCTGCTGGAAGTCC 831
DB 832 TTGTGACTTCAAGAGGTGCCCCCGTGGAGGCGCAATCTCACTTACCTGCTGGAAGTCC 831
QY 644 CGGGGTGGCGCAGTTCGAGACCGCGTGTGCAAGACCAACCGCGTGAAGGCTTTGGG 891
DB 892 CGAGTGTGTGCACCAAGATCATGAGGAGCGCAATCTCAGTTCATCTCAGTTCAGTTCAG 891
QY 704 GGGGGGAGAGAGAGACCTCTCTGCGCTGGGCTGGGAGCGAAGCGAAGCGGAGGAG 951
DB 952 GGGGGGAGAGAGAGACCTCTCTGCGCTGGGCTGGGAGCGAAGCGGAGGAGGAG 951
QY 764 TACCTGTGTGGAAGGCGCAGTGTGCAAGGTCCTCTCAATCAAGCAAGAGTGTGGAAG 1011
DB 1012 TACCTGTGTGGAAGGCGCAGTGTGCAAGGTCCTCTCAATCAAGCAAGAGTGTGGAAG 1011
QY 824 GTATGAGGAAGGCGCTGTGCGCATTTACCTCACTGAGGATGAGGAGACTTGTCTC 1071
DB 1072 GTGCTGAGGAAGGCGCTGTGCGCATTTACCTCACTGAGGATGAGGAGACTTGTCTC 883
QY 884 AGCATGTGTGCGCAGGCTGTGCAATCTGAGGAGCAATCTGCTGCTGAGGAGGAGGAG 1131
DB 1132 AGCATGTGTGCGCAGGCTGTGCAATCTGAGGAGCAATCTGCTGCTGAGGAGGAGGAG 1131
QY 944 AATGCCAAGGTTACTTCACTGAGAACCAAGCTCAATTTCTGACAGGCTCTTGTGTGGA 1191
DB 1192 AATGCCAAGGTTACTTCACTGAGAACCAAGCTCAATTTCTGACAGGCTCTTGTGTGGA 1191
QY 1004 GGTGCACACTTGTAGGGAAGCGCTGACCCAGCAAGAAATCAATCCCAAGGGGAGAGGAG 1251
DB 1252 GGTGCACACTTGTAGGGAAGCGCTGACCCAGCAAGAAATCAATCCCAAGGGGAGAGGAG 1251
QY 1064 CTGAGCCCACTCAATCACTTGTAGGGAAGCGCTGACCCAGCAAGAAATCAATCCCAAGGGGAGAGGAG 1311
DB 1312 CTGAGCCCACTCAATCACTTGTAGGGAAGCGCTGACCCAGCAAGAAATCAATCCCAAGGGGAGAGGAG 1311
QY 1124 TACAGCGGCAATTCACCTGCTGTGCAAGAAATCAATTTAGTACTGAGGCTTGAAGGAG 1371
DB 1372 TACAGCGGCAATTCACCTGCTGTGCAAGAAATCAATTTAGTACTGAGGCTTGAAGGAG 1371
QY 1184 GGTGAGGCGCCAGGCTGTGCAAGAAATCAATTTAGTACTGAGGCTTGAAGGAG 1431
DB 1432 GGTGAGGCGCCAGGCTGTGCAAGAAATCAATTTAGTACTGAGGCTTGAAGGAG 1431
QY 1244 GAAGTGTTCAGATTAAGAGCTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1491
DB 1492 GAAGTGTTCAGATTAAGAGCTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1491
QY 1304 CAGAGCTTTCATGAGAGCTGCTCAAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1551
DB 1552 CAGAGCTTTCATGAGAGCTGCTCAAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1551
QY 1364 ATGCGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1611
DB 1612 ATGCGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1611
QY 1424 AAGTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1423
DB 1472 AAGTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1471
QY 1484 GAGCTGAGCTTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1483
DB 1572 GAGCTGAGCTTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1571
QY 1544 CACAAGCTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1791
DB 1792 CACAAGCTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1791
QY 1604 CATTAATGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1603
DB 1852 CACTATGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1851

```

QY 1664 CTCTTCCGGAACTGAGAGACCATGTGTCACCTCATATGAACCCCATCATATGAGCCCACTGC 1723  
DB 1912 CTCTTCCGGAACTTAAAGAGACCATGTGTACTCAAGAAATCCATTATATGAGCCAGTGC 1971  
QY 1724 TTTTGACAAAGATGAGTACTAGTATACAAGAACGGCCAG-GAGCGTGGCCACCCAGTTTCAAG 1782  
DB 1972 TTTGACCGGAGGAGCTAGTATACAAGAACGGCCAGAGACGCTGGCCACCCAGTTTCAAG 2031  
QY 1783 ATGACCTCTCTGACCTGTGTGAGATCTGAGTCTAAAGAGACCTGCTCATATGCGGTGC 1842  
DB 2032 ATGACCTCTCTGACCTGTGTGAGATCTGAGTCTAAAGAGACCTGCTCATATGCGGTGC 2091  
QY 1843 ATCAAGCCAAAGACGACCAAGACGCGGGTGCCTTTGATGAGTGTCTCATCCAGATCAG 1902  
DB 2092 ATCAAGCCAAAGATGATGCCAAAGACGCGGGCTTTGAGAGAGTGTGTATCCGCCACAG 2151  
QY 1903 GTGAAGTACTGTGGACTGTATGAGAAATCTGCGCGTGGCCAGAGCTGGCTTTCCCTATGCT 1962  
DB 2152 GTGAAGTACTGTGGGCTGTGTGAAACCTGCGCGTGGCCAGAGCGGGCTTTCCCTATGCT 2211  
QY 1963 GCGAATATGAGGCTTTCCTGAGAGTCAAGTCACTGTGCCAGAGACATGSGCCATG 2022  
DB 2212 GCGAATATGAGAGCTTTCCTGCAAGATCAAGTCACTGTGCCAGAGAGTGTGGCCACG 2271  
QY 2023 TGGGAGAGACGCGCCAGGATGTGTGGCCGTGTGTGTCAGACACTCGGCTACAAGCCA 2082  
DB 2272 TGGGAGAGACGCGCCAGGATGTGGGCTGTGTGTGTCAGACACTCGGCTACAAGCCA 2331  
QY 2083 GAAAGTACAAATATGGGACGAGACTAAGATCTTATCCGATTTCCCAAGACCTTATTTGCC 2142  
DB 2332 GAAAGTACAAATATGGGACGAGCAAGATCTTATCCGATTTCCCAAGACCTTATTTGCC 2391  
QY 2143 ACAGAGGACTCCCTGGAAATCCGGGGGAGAGTACAGCACCAGAGTCCAGCGGCTGC 2202  
DB 2392 ACAGAGGATGCTCTGAGGTCCGGGGGAGAGCTGTGCCACAGAGTCCAGAGCTGTGC 2451  
QY 2203 AAGGCTTTTCATTTGGGAGCAAAATTTTCGCGGTGAAGCAGTACAGCCATCTGTATCCAG 2262  
DB 2452 AAGGCTTTTCATTTGGGAGCAAAATTTTCGCGGTGAAGATCAGCCATCTGTATCCAG 2511  
QY 2263 TCATGTGCTGGGACACTGTGGCCGGAGGAAGGACCAAGAGAGTGGGACGCCAG 2322  
DB 2512 TCGTGTGTGCTGTGAACACTGTGGCCGGAGGAAGGACCAAGAGAGTGGGACGCCAG 2571  
QY 2323 ACCATCCGTCGACTATCCGTGGCTTTCATTTTGGGCAATTCACCCCGGTGCTGGAAT 2382  
DB 2572 ACCATCCGTCGACTATCCGTGGCTTTCATTTTGGGCAATTCACCCCGGTGCTGGAAT 2631  
QY 2383 GCTTCTTCTTGGACCACTGTGGCGCTTCATTTTGGCTTAACCTGAGGCGGCAACTGCC 2442  
DB 2632 GCTTCTTCTTGGACCACTGTGGCGCTTCATTTTGGCTTAACCTGAGGCGGCAACTGCC 2691  
QY 2443 CGGAATGTTCTGGACACTGTGGCCACACCCCACTGCTGCTGAGAGAGCCCTCAGAA 2502  
DB 2692 CGGAATGTTCTGGACACTGTGGCCACACCCCACTGCTGCTGAGAGAGCCCTCAGAG 2751  
QY 2503 CTGCTACGGAAGTGTGATGGAAGACATGTGTGAATACTGCTGGGAGCATCAGCCCT 2562  
DB 2752 CTGCTACGGAAGTGTGATGGAAGACATGTGTGAATACTGCTGGGAGTATCAGCCCT 2811  
QY 2563 GAGTGAAGCAGCAGCTGACGCAAAAGCGGTGTGATGAAATTTTCAAGGGCAGAGAG 2622  
DB 2812 GAGTGAAGCAGCAGCTGACGCAAAAGCGGTGTGATGAAATTTTCAAGGGCAGAGAG 2871  
QY 2623 GACAACTACCCAGAGTGTCCCAAGACTTTCATTTAGCACAGCGCTGGCACAGAGAG 2682  
DB 2872 GATTAATTAATCTCAGAGTGTACCAAGGCTTTCATCAGACACTCGGCTGTGTAGATGAG 2931  
QY 2683 ATCAGCCCAAGAGTGTCTCAATCTTGGGCTGTGAACCCATCAGATACCGTCCCGTG 2742  
DB 2932 ATCAGCCCAAGAGTGTCTGAGGCTTGGGCTGTGAACCCATCAGATACCGTCCCGTG 2991

QY 2743 GTAATAATACGACCTGAAGGTTTACAAGCCTGCCCCCGGACCTGCTGTACGCCAGT 2802  
DB 2992 GTGAATATACGACCTGAAGGTTTACAAGCCTGCCCCCGGACCTGCTGTGTACGCCCAAC 3051  
QY 2803 GCTGTGTATGTGTGAGATGCTTAAAGTCAAGCAGAGAAATTTATTTATGCCAATACC 2862  
DB 3052 GCGGTGTATGTGTGAGATGCTTAAAGTCAAGCAGAGAAATTTATTTATGCCAATACC 3111  
QY 2863 GGAATCTCTGTAGTATGCTGTAGTATACCTATTTTGTGCTTACAGTGTGAGTGAAGC 2922  
DB 3112 GGAATCTCTGTAGTATGCTGTAGTATACCTATTTTGTGCTTACAGTGTGAGTGAAGC 3171  
QY 2923 AACAGCAGAGGAGGATGTGTGTGTCAGAGTGTATCATGTATGACACTAACAAG 2982  
DB 3172 AATAAGCAAAAGGAGATGTGTGTGTCAGAGTGTACCAAGTATGAGACGCTGACCAAG 3231  
QY 2983 ACGGCTCTAGTGTACCGGCTGAAACATATCAACATTAACCAAGGCGCATTAAGCTT 3042  
DB 3232 ACGGCTCTAGTGTACCGGCTGAAACATATCAACATTAACCAAGGCGCATTAAGCTT 3291  
QY 3043 GCAGGGGTCAGAGCAGGAGCATATGACTTATGACTTACATCGGGCTCAGAGCTTTCATC 3102  
DB 3292 GCAGGGGTCAGAGCAGGAGCATATGACTTATGACTTACATCGGGCTCAGAGCTTTCATC 3351  
QY 3103 ACCAAGGCTAAGATGCCCACCTGTGTGTGTCGCCCAAGGCTGAATTTCTGGTATGA 3162  
DB 3352 ACCAAGGCTAAGATGCCCACCTGTGTGTGTCGCCCAAGGCTGAATTTCTGGTATGA 3411  
QY 3163 ACGCTGTGTGTGACGCTCTGTACTGTGTATGCTTCCCTTACTCCCTCTCCCTCGCA 3222  
DB 3412 ACGCTGTGTGTGACGCTCTGTACTGTGTATGCTTCCCTTACTCCCTCTCCCTCGCA 3471  
QY 3223 CTAAACAAAGCTCAAGCTTCCAAACAGGAGTCAATGACACCTTCAAAACACAGCTGC 3282  
DB 3472 GTTACCAAGAGCTCCAGACTTCCAGACAGGAGCACCAGGGACACCCGAAAGCCAC - CTGC 3529  
QY 3283 AAATCTCTCTTGTGTGTGTCGCCCTCTTGAAGTGTATCAG - GAGCCAGGAGCTACCC 3340  
DB 3530 AAATCTCTCTTGTGTGTGTCGCCCTCTTGTGAAGGAGCAGACAGGGGCTGAGAGCTACCC 3589  
QY 3341 ATGATGTGGGTCAGGCGGCGCACCAATATGAAGAGCAGAGGCTGTAGAGGCGCAG 3400  
DB 3590 AAGAGTGTGGTCAGGCGGCGCACCAATATGAAGAGCAGAGGCGCAGAGGCGCATGCA 3649  
QY 3401 GCCCT-CTGCTGATGCAATATATATAGACAAAGGAATTTAATCTGAGGTTTTCTCTGA - 3458  
DB 3650 GCCCTGATGCAATATATATAGAGAGGAGAACTTTGTGTAGGTTTTCTCTGAG 3709  
QY 3459 GATTTTGTATGCTTATATAGAAACTATTTTAAAGAAAGCCATTTTCTACCCCTAAG 3518  
DB 3710 GATTTTGTATGCTTATATAGAAACTATTTTAAAGAAAGCCATTTTCTACCCCTAAG 3769  
QY 3519 ACATGTGATGT 3578  
DB 3770 ACAGTGTGATGTG-TTTTCCCTGACTCCAGAGGCAAGGAATGTATACCAAGAGTGTGT 3828  
QY 3579 GGGGTGGGCTGGAAGTGTCTTCTTGTGCAACCCCTTCTTATTTCCCTGTGCTGCTG 3638  
DB 3829 GGGGTGGGCTGT 3888  
QY 3639 CATCC 3643  
DB 3889 CTTCG 3893

RESULT 2  
AA574358  
ID AA574358 standard; cDNA; 3392 BP.  
XX AA574358;  
AC  
XX  
DT 13-FEB-2002 (first entry)  
XX

DE DNA encoding novel human diagnostic protein #10162.  
 KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 OS Homo sapiens.  
 PN WO200175067-A2.  
 PD 11-OCT-2001.  
 PF 30-MAR-2001; 2001MO-US08631.  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 PA (HYSB-) HYSBQ INC.  
 PT Drmanac RT, Liu C, Yang YT;  
 DR WPI: 2001-639362/73.  
 XX P-PSDB: ABC10171.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT in biodiversity -  
 PS Claim 1: SEQ ID NO 10162; 103bp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC and gene mapping, and in recombinant production of (II). The  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement (II) and its binding partners are useful in medical  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations in  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC diagnostic coding sequences. A564197-A594564 represent novel human  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.int/pub/published\_pcl\_sequences.  
 XX  
 SQ Sequence 3992 BP: 775 A; 993 C; 999 G; 625 T; 0 other:  
 Query Match 69.6%; Score 2549; DB 23; Length 3392;  
 Best Local Similarly 86.5%; Pred. No. 0;  
 Matches 2906; Conservative 0; Mismatches 440; Indels 13; Gaps 8;  
 QY 42 CGGCCCTGGCAGTGCAGGGGTCGAGTACCATGAGAGCGCCCTGACTGCCGAGACC 101  
 DB 34 CGGCCCTGGCAGTGCAGGGGTCGAGTACCATGAGAGCGCCCTGACTGCCGAGACC 101  
 QY 102 GGGTAGGGGTCAGGACTTGTCTGCTGAGAAATTCACAGTGAAGGCTGCTTCATG 93  
 DB 94 GGGTAGGGGTCAGGACTTGTCTGCTGAGAAATTCACAGTGAAGGCTGCTTCATG 161  
 QY 162 AGAACCTCGGGGCGGCTTCGGGGAACCTCATTTATACCTACATCGCTCTGCTGAG 221  
 DB 154 AGAACCTCGGGGCGGCTTCGGGGAACCTCATTTATACCTACATCGCTCTGCTGAG 221  
 QY 222 TCTCTGTAATTCCTACGAGACCTACAGATCTACAGCGGAGATATGGAACGCTACG 213  
 DB 214 TCTCTGTAATTCCTACGAGACCTACAGATCTACAGCGGAGATATGGAACGCTACG 281  
 QY 282 GTGTCGTCAGTTTCTATGAAATACACATCTTTGTCGATGCTGACATGTAATAC 341  
 DB 274 GTGTCGTCAGTTTCTATGAAATACACATCTTTGTCGATGCTGACATGTAATAC 341  
 QY 342 GGGCAGCTTGTACTGAGCGCTCGGAGACAGCAGTGAATGATTTGAGAGAGTGGGAG 333  
 DB 334 GAGCAGCTGCGACAGGAGCGCTCGGAGACAGCAGTGAATGATTTGAGAGAGTGGGAG 401  
 QY 402 GCAAGACAGAGCCACCAAGAACTGCTCCAGTCTATGAGAGAGCTGCCAGCCCTG 393  
 DB 394 GCAAGACAGAGCCACCAAGAACTGCTCCAGTCTATGAGAGAGCTGCCAGCCCTG 461  
 QY 462 AACGGGTCGGGCGAGTGGAGACCGCTGTGTCAGAGCAACCCGCTGTTAGAGCTTTG 453  
 DB 454 AACGGGTCGGGCGAGTGGAGACCGCTGTGTCAGAGCAACCCGCTGTTAGAGCTTTG 521  
 QY 522 GGAATGCCAAGCTCTCCGCAAGATTAATCCAGCGGTTTGAAGATGATGATGTC 581  
 DB 514 GAAATGCCAAGACCTCTCCGCAAGATTAATCCAGCGGTTTGAAGATGATGATGTC 581  
 QY 582 AGTTGACTTCAAGGGTCCCGCTGGAGAGCCACATTTCTCAGTTACTCTGGAAGAGT 573  
 DB 574 AGTTGACTTCAAGGGTCCCGCTGGAGAGCCACATTTCTCAGTTACTCTGGAAGAGT 641  
 QY 642 CCGGGTCGTGCACCAAAATACAGAGAGCGAATCTTCAAGCTTTTACAGCTACG 633  
 DB 634 CAGCAAGTGGTGACCAAGAAATCATGAGAGCGAATCTTCAAGCTTTTACAGCTACG 701  
 QY 702 AAGGGGGGAGAGAGAGACTTCCTGCTGCTGGCTGGCTTGAAGAGAACCCGAGCTACT 693  
 DB 694 AAGGGGGGAGAGAGAGACTTCCTGCTGCTGGCTGGCTTGAAGAGAACCCGAGCTACT 761  
 QY 762 TGTACTCTGTAAGAGGCGCACTGTGTCAGAGTCTCTTCATCAACAGAGAGTACTGA 753  
 DB 754 TGTACTCTGTAAGAGGCGCACTGTGTCAGAGTCTCTTCATCAACAGAGAGTACTGA 821  
 QY 822 AGTTTATGAGAGAGAGGCGCTGCTGCTGATTTGACTTCACTGAGATGAGAGAGCTTGC 813  
 DB 814 AGTTTATGAGAGAGAGGCGCTGCTGCTGATTTGACTTCACTGAGATGAGAGAGCTTGC 881  
 QY 882 TCAGCATCTGTGCGAGCGCTCTCATATCTGGGCAACATCCATTTGCTGTGACGAGACA 873  
 DB 874 TCAGCATCTGTGCGAGCGCTCTCATATCTGGGCAACATCCATTTGCTGTGACGAGACA 941  
 QY 942 GCAATGCGGAGGCTGAGTCTTATTTGAGCAACATCTTCTGCTGCAACAGAGACA 933  
 DB 934 GCAATGCGGAGGCTGAGTCTTATTTGAGCAACATCTTCTGCTGCAACAGAGACA 1001  
 QY 1002 AAGGTATCAACACTTAGGAGAGCGCTGACCCAGAGAGATGATCGCCAGAGGAGAGAC 993  
 DB 994 AAGGTATCAACACTTAGGAGAGCGCTGACCCAGAGAGATGATCGCCAGAGGAGAGAC 1061  
 QY 1062 TCTGAGCCGCTGAGCACTTGAACAGCGGATATGCAAGGATGGGCTTGCAGAGCTG 1053  
 DB 1054 TCTGAGCCGCTGAGCACTTGAACAGCGGATATGCAAGGATGGGCTTGCAGAGCTG 1121  
 QY 1122 TGTACAGCGGAGATTAACCTGCTGCTGTCAGAAAGATCAATAGTCTACTGAGCTTAAG 1113  
 DB 1114 TGTACAGCGGAGATTAACCTGCTGCTGTCAGAAAGATCAATAGTCTACTGAGCTTAAG 1181  
 QY 1182 AGCGTGAAGGCCAGCTGAGTGGGAGAACCCAGCGCTTGGGCTCTGAGATTTATGAGCT 1173  
 DB 1174 AGCGTGAAGGCCAGCTGAGTGGGAGAACCCAGCGCTTGGGCTCTGAGATTTATGAGCT 1241  
 QY 1242 TTGAAGTGTTCAGCAATACAGCTTTCAGCACTTGTGATCAATCTGCAATGAGAGAC 1233  
 DB 1234 TTGAAGTGTTCAGCAATACAGCTTTCAGCACTTGTGATCAATCTGCAATGAGAGAC 1301  
 QY 1302 TGCAGCAGCTCTTCACTGAGCTGCAAGTGGAGAGAGAGATATGAGAGCTGAGG 1293  
 DB 1294 TGCAGCAGCTCTTCACTGAGCTGCAAGTGGAGAGAGAGATATGAGAGCTGAGG 1361  
 1353



OY	1362	GCATGCGGTGGGAAACCTGTCCACACTTCTTCAACAACAAGTCACTGTGTACCTGGTAGAGC	1421
Db	1354	GCATCGCGGTGGGAAACCCGTCCACTATTTTCAACAACAATAATCATCTGTATCTGGTAGAGG	1413
OY	1422	AGAGATTCAAGGGCATTCAATCTCCATCTTGTGATGAAGAGTGGCGGTCTGGGAGAGCCA	1481
Db	1414	AGAAAGTTTAAAGGCATCATCTGCATTTTGGATGAGSAGTGTCTGGCCCGGGGGAGAGCCA	1473
OY	1482	CGGACCTGACCTTTCTTGAGAGAAAGTTGGAGAGACATGTCAAGCCCCACCTCATTTCTGA	1541
Db	1474	CAGACCTGACCTTCTCTGGAGAAAGCTGGAGGATATCTGTCAAGCAACCATTCACACTTCTGA	1533
OY	1542	CGCACAAGCTCGCTGACACAGAACCAAGSAAATCCCTAGACCAGAGGAGTTCCGCTTC	1601
Db	1534	CGCACAAGCTGGCTGACACAGAGCACCAGSAAATCTCTGGGGCCGAGAGGGAAATTTCCGCTTC	1593
OY	1602	TGCATTTATGCGGGAGAGGTACCTACAGATGTGACGTGGGTTTCTGGATPMAAACATGACC	1661
Db	1594	TGCATTTATGCGGGAGAGGTACCTACAGGTGACGTGACCGGTTTCTGGACMAAACAATGACC	1653
OY	1662	TCTCTTTCCGGAAACCTGAAGAGACCATGTGACAGTCAATGAACCCCATCATATGGCCAGT	1721
Db	1654	TTTCTTTCCGGAAACCTTAAAGAGACCATGTGTAGTCAAGAATACTTATGAGCCAGT	1713
OY	1722	GCTTTGACAAAGTAGTGAAGCTAGTGACAAAGAGCGCCAG-GACGCTGGCCACCACTTCA	1780
Db	1714	GCTTTGACCCGAGAGGACACTAGTGACAAAGAGCGGCGCAGAGCGGTGGCCACCACTTCA	1773
OY	1781	AGATGAGCCTCTCCACCTGTGTGAGATCCTGTGAGTCTTAAGAGCCTGCTATATTCGGT	1840
Db	1774	AGATGAGCCTCTCTCACCTGTGTGAGATCCTGTGAGTCTTAAGAGCCCGCTTACGTCCCT	1833
OY	1841	GCATCAAGCCAAAGCAGCGCCACAGCCGCGGTGCTTTGATGAGGTGCTATCCGACATC	1900
Db	1834	GCATCAAAACCAATGATGATGCCAAAGCCGCGGCTTTGACAGAGGTGCTATCCGCGCAC	1893
OY	1901	AGGTGAAGTACTCTGGACATATGAGAGATCTGCGCTGCGCAGAGCTGGCTTTGGCTATC	1960
Db	1894	AGGTGAAGTACTCTGGGCTTTGTGGAAACCTGTGTGTGCGCAGAGACTGGCTTTGGCTATC	1953
OY	1961	GTGCGAAATATGAGGCGTTTCTGTGAGAGGTACAGTCACTGTGSCCAGAGCATATGGCCA	2020
Db	1954	GCCGCAATATGAAAGCTTCTCTCCAAAGGTACAGTCACTGTGSCCAGAGCATATGGGCCA	2013
OY	2021	TGTGGCGAGGACGGCCCGCAGAGATGTGTGGCCCTGTGTGTGACAGACCTCGGCTACAAGC	2080
Db	2014	CGTGGCGAGGACGGCCCGCAGAGATGGGGTGGCTGTCTGTGTCGACACACTCGGCTACAAGC	2073
OY	2081	CAGAAAGTACAAATATGGCGAGAGCTAAGATCTTCAATCCGATTTCCCAAGACCTATTATG	2140
Db	2074	CAGAAAGTATCAAAATATGGCGAGAGCAAGATCTTCAATCCGATTTCCCAAGACCTATTATG	2133
OY	2141	CCACAGAGGACTCCCTGTGAAGTCCGGGGGAGAGTGTATCCACCAAAATATCCAGGGGGCT	2200
Db	2134	CCACAGAGGATGCCCTGTGAAGTCCGGGGGAGAGTGTATCCACCAAAAGATATCCAGGTGCT	2193
OY	2201	GGAGGGGCTTTCAATTTGGCGAGAGAAATTTTCTCCGGGTGAAGGATCAGCAATCTGTATCC	2260
Db	2194	GGAGGGGCTTTCACTTGGCGGAGAAATTTCTCCGGGTGAAGAGATCAGCAATCTGTATCC	2253
OY	2261	AGTATGTGTGGCGTGGCACACTGGGGCGGAGAGGACCAAGAGAGAAATGTGGGGCAGC	2320
Db	2254	AGTGTGTGTGGCGTGGGAAACACTGGGGCGGAGAGGACCAAGAGAGAAATGTGGGGCAGC	2313
OY	2321	AGACATCCGTCGACATCAATCCGTGGGCTCAATTTTGGGCAATTCACCCGGGTGCTGTGA	2380
Db	2314	AGACATCCGCGGCTCATCCAGAGGCTCAATCTGTGGCCACGCCCCCGCTGCCCCGAGA	2373
OY	2381	ATGCGCTTTCTTTGAGCACGACGTGGCGCTCATTTTGTCTTAACCTGAGGCGGCAACTGC	2440
Db	2374	ATGCGCTTTCTTTGGACCATGTGTGGCGACGCTTTTGTGTAAACCTGAGGCGGCAAGCTGC	2433
OY	2441	CCCGGAATGTCT-GGACACCTCTCTGGCCACACCCCAACCTGCTCCCTGAGAGAGGCTTCA	2499

Dp	2434	CCCGGAATGTCCTGCGACACCTGCTGGCCACGCCCCCACTGCTCGGTAGAGGCTCTCA	2493
Qy	2500	GAACTCGTACGGAGCAACTGTGCATGAAGAACATGGTGTGAAGTACTGCCGAGCATCAGC	2559
Dp	2494	GAGCTTCTCGGGGGAGTGTGTGCATTAAGAACATGTGTGGAAATACTGCCGGAGTATCAGC	2555
Qy	2560	CTGTAGTGGAGCAGCAGCTGTCAGCAAAAGGGCGGTGGATGTGAATTTTCAAGGGCAAG	2619
Dp	2554	CTTGAGTGGAAAGCAGCATCTGTCAGCAGAAAGGGCGGTGGATGTGAAGTCTTTCAAGGGCAAG	2613
Qy	2620	AAGGACAACTACCCCCGAGTGTGCCAGTCTTCATTAAGACACAGCGCTTGGACAGAG	2679
Dp	2614	AAGGATTAATTACCTCTGAGTGTATCCCAAGGCTCTTATCAGACACTGCGGTGGTACAGAT	2677
Qy	2680	GAGATACAGCCCCAGAGTGTCTCAATCTCTTGGGCTTGAAGCCCATCAGATATGCCGTGCC	2733
Dp	2674	GAGATCAACCCCCGAGTGTCTGAGCGCCTTGGGCTCTGAGACCCATTCAGTATGCTGTGCC	2733
Qy	2740	GTGTGTAATAATACGACCGTAAAGGGTTTACAGACCTC -GCCCGCGGACGCTGTCT - -CA	2799
Dp	2734	GTGTGTAATAATACGACCGTAAAGGGTGTACAAACCTCGGTCCCGGACGCTGTCTGTACAGG	2799
Qy	2795	CGCCCACTGCTGTGCTATTTGTGGAGGATCTTAAG -TCAGAGAGAAATGTATATGCC	2855
Dp	2794	CCCAACGCGCGTGTATGTATGTGGAGAGCCCAAGAGTTCAAGAGAGGATGTATATACGCC	2855
Qy	2854	-AACCTAACCGGAATCTCTGTACAGTACGATGAGTGAATAGCTTTTGTCTTCACGTGCA	2913
Dp	2854	AAACCTGACCGGGAATCTCTGTACAGACGCTGACAGCAGACGCTTTTGTGCTTCATGTACA	2913
Qy	2913	GCGTGAAGACACAAGCAGAGAGGAGATGTGTGCTGTGCAGAGTATCATGTGATCGAGAC	2977
Dp	2914	GCGTGCGGACAAATTAAGCAAAAAGGAGATGTGTGCTGTGCAGAGTATCATGTGATCGAGAC	2977
Qy	2973	ACTTAACCAAGACGGCCCTCACTAGTGTGACCGCGGTGAACAATATCAATCAACCAAGGCGAG	3033
Dp	2974	GCTGACCAAGACAGCCCTCACTAGTGTGACCGCGGTGAACAATATCAATCAACCAAGGCGAG	3033
Qy	3033	CATTAACGTTTTCAGGGGGGTCCAGCAGGACGAGCATCATTTGACTTACATTCGCGCTCACA	3097
Dp	3034	CATCAACGTTTTCAGGGGGGTCCAGCAGGACGAGCATCATTTGACTTACATTCGCGCTCACA	3097
Qy	3093	GCTTCTCATCCACAAGGCTAAGAAATGGCCACCTGGCTGTGTGTGGTGGCCCAAGGCTGAATTC	3157
Dp	3094	GCTGCTCATCCACAAGGCTAAGAAATGGCCACCTGGCTGTGTGTGGTGGCCCAAGGCTGAATTC	3157
Qy	3153	TCGGTGTATGAAGGCTGGCGGTGGACCGTCTGACTGCTGATGCTTCCCTTAATGTCCTCTCC	3217
Dp	3154	TCGGTGTATGAAGGCGCCACATGTGCACATCCCAAGGCCCAAGGCTTGTCTTCTCTCTCT	3217
Qy	3213	TCGCCCTCGAGTTTACCAAAAACCTCAACCTTCCCAACAGGGATCATGAGACACCTCAAAA	3277
Dp	3214	CCCTTTCCTCAAGTTTACCAAAAGTGTGAATTTTCCAGACAGGAGACCCAGAGACACC -CCGAA	3277
Qy	3273	CCCAAGCTCAAACTCCTGCTTCTGCTGTGCGCCCTCTTTTGAAGGTATGTACAG -GAGCCAGG	3337
Dp	3272	GCCCAAGCTCAAAATTTTCCACCTCTCTGCGCATCCCTTTCTTGAAGGAGCAGCAGAGGGCCAG	3337
Qy	3331	GAGCTACCCCATGTAGTGGGCGCAGGCGCGGCGACACCAATATGAAAAGCAGAGGCTGTAGC	3389
Dp	3332	GAGCTACCCCATGTAGTGGGCGCAGGCGCGGCGACACCAATATGAAAAGCAGAGGCGCTAGAGC	3390
RESULT 3			
AA574359			
ID	AA574359 standard; cDNA; 3771 BP.		
XX	AA574359;		
XX	13-FEB-2002 (first entry)		
DE	DNA encoding novel human diagnostic protein #10163.		

XX XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.  
KW Homo sapiens.  
OS  
PM WO200175067-A2.  
PD  
XX  
PF 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YF;  
DR WPI: 2001-639362/73.  
XX P-PSTDB: ABG10172.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
XX biodiversity

Claim 1: SEQ ID No 10163; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics of (II). The  
CC for identifying expressed genes. (I) is useful as expressed sequence tags  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. Ab564197-AM594564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pcl\_sequences.

Sequence 1171 pp. 1-10

```

Query Match      40.8%  Score 1492;  DB 23;  Length 3771;
Best Local Similarity 81.7%;  Pred. No. 0;
Matches 1847;  Conservative 0;  Mismatches 320;  Indels 93;  Gaps
7.

Qy  591 TCAAGGTCCTCCCTCGGAGGCGACATTCCTCAGTTACTCCTGGAAAGTCCCGGGTG 650
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  242 TGAAGGTCCTCCCTCGGAGGCGACATTCCTCAGTTACTCCTGGAAAGTCCCGGGTG 650
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy  651 TGCACCAAAATCAGCGAGAGCGGAACCTTCACGCTTTTATCAGCTCTGTGAGGGGGCG 301
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  302 TGCACCAAAATCAGTGGGAGCGGAACCTTCACGCTTTTATCAGCTCTGTGAGGGGGCG 301
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy  711 AGGAGGAGACCTTCCTCGGCTTGGGCTTGGAAAGGAAACCCCGAGAGCTACTGTACCTGG 770
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  362 AGGAGGAGACCTTCCTCGGCTTGGGCTTGGAAAGGAAACCCCGAGAGCTACTGTACCTGG 770
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy  771 TGAAGGCCAAGTGTGCCAAAGCTTCCTCCATTCACAGCAAGAAGTACATCGGAAGGTTATGA 830
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  422 TGAAGGCCAAGTGTGCCAAAGCTTCCTCCATTCACAGCAAGAAGTACATCGGAAGGTTATGA 830
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

[illegible]

```

Db 1556 TCCCTACCCCGAGCCGCTTTGACGAGGTGCTGATCCGCCACAGTGAACCCGCTT 1615
Qy 1953 TGGCTATCGCGCAAAATATGAGGCTTTCGCGAGAGGTACAACTGCTGCCACAGAC 2012
Db 1616 TGGCTATCGCGCAAAATATGAGGCTTTCGCGAGAGGTACAACTGCTGCCACAGAC 1675
Qy 2013 ATGGCCATGTGGGAGAGAGGCGCCAGAGATGTGTGGCCGTGTTGAGTACAGCTCGG 2072
Db 1676 GTGGCCAGATGGGAGAGAGGCGCCAGAGATGGGCTGTGCTGTGCTCCGACACTGGG 1735
Qy 2073 CTACAAAGCCAGAGAGTACAAATGGGAGAGTAAAGATCTTCATCCGATTCGCCAAGAC 2132
Db 1736 CTACAAAGCCAGAGAGTACAAAGATGGGAGAGACCAAGATCTTCATCCGATTCGCCAAGAC 1795
Qy 2133 CTATATTTGGCAGAGAGACTCCCTGGAGTGGGCGGAGAGCTGAGTCCAGCAAGATCCA 2192
Db 1796 CCGTTTGGCAGAGAGATGCTGGAGGTGCGGCGGAGAGCT----- 1840
Qy 2193 GGGGCGCTGGAGGGGCTTTCATTTGGCGACAGAAATTTCTCGGGTGAAGCGATCAGCAT 2252
Db 1841 -----GGCCAT 1846
Qy 2253 CTGTATTCAGTATGTGGCTGGCGACACTGGCCGAGAGAGAGCAAGAGAACTG 2312
Db 1847 CTGTATTCAGTATGTGGCTGGCGACACTGGCGGAGAGAGCAAGAGAACTG 1906
Qy 2313 GGGAGCCAGAGACTCGGTGACTCGGTGCTTCATTTTGGCCATTCACCCGGTG 2372
Db 1907 GGGGAGCAGACCATCGGGGCTCATCCGAGGCTTCATCTCCGACACCCCGCGCT 1966
Qy 2373 CCCTGAGATGCTCTTCTTTGGACCACTGCGGCTCATTTTGTGCTTAACCTGAGCG 2432
Db 1967 CCGGAGAACGCTTCTTCTTGAGACATGCGGACCTCTTTTGTGTAACCTGAGCG 2026
Qy 2433 GCACTGCCCCGAGATGTTCTGAGACCTCTGCCCCACACCCCTGCTGAGAGA 2492
Db 2027 GCAGCTGCCCCAGAAATGCTCTGAGACACTGTCGCCACCCCTGCTGCTGCTGA 2086
Qy 2493 GGGCTGAGAACTGCTGAGGAGTGCATGAGAAACATGATGGAGTACGCGGAG 2552
Db 2087 GGGCTGAGAACTGCTGAGGAGTGTGCAATAGAAACATGATGGAGTACGCGGAG 2146
Qy 2553 CATCAGCCCTGAGTGAAGAGCAGCTGACAGAAAGGCGGTGCTAGTAAATTTTCAA 2612
Db 2147 TATCAGCCCTGAGTGAAGAGCAGCTGACAGAAAGGCGGTGCTAGTAAATTTTCAA 2206
Qy 2613 GGGCAGAAAGAGCACTACCCCGAGAGTGTCCCGACACTTTATTAAGCACAGGCTTG 2672
Db 2207 GGGCAGAAAGAGTAAATTAATCCCTGAGAGTACCCAGGCTTTCATCAGCACTGCGGCTTG 2266
Qy 2673 CACAGAGAGATAGCCCGCAGAGTGTCTCAATGCTTGGGGCTGGAACCCATCCAGTATGC 2732
Db 2267 TACAGATGAGATAGCCCGCAGAGTGTCTGAGAGGCTTGGGCTTGAAGCCATTAAGTATGC 2326
Qy 2733 CGTGCCCTGTGTAATATGACACCGTAAAGGTTACAAAGCTGCGCCCGAGCTGTGCT 2792
Db 2327 GGTGCTGTGTAATATGACACCGCAAGGCTACAAAGCTGCGCCCGAGCTGTGCT 2386
Qy 2793 CACGCCAGTGTGTGTCTATTTGTGAGAGATGCTAAAGTC 2832
Db 2387 CACGCCAGGCGGTGTCTATCTGTGAGAGAGCTCAAAAGTC 2426

```

RESULT 4  
AA160933  
ID AA160933 standard; cDNA: 1727 BP.

AC AA160933;

DT 22-OCT-2001 (first entry)

DE Human polynucleotide SEQ ID NO 4922.

```

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX Homo sapiens.
PN WO200153312-A1.
PD 26-JUL-2001.
PF 26-DEC-2000; 2000MO-US34263.
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX (HYSE-) HYSEQ INC.
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Zhang J, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX MPI: 2001-442253/47.
XX P-PSDB; AAM41777.
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries.
XX Claim 1: SEQ ID NO 4922; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SO Sequence 1727 BP; 364 A; 500 C; 543 G; 320 T; 0 other;
XX
Query Match 31.3%; Score 1144.8; DB 22; Length 1727;
Best Local Similarity 88.2%; Pred. No. 1,8e-312;
Matches 1267; Conservative 0; Mismatches 167; Indels 2; Gaps 2;
XX
Qy 44 GCCCTGGCAGTGCAGGCGGTTCGATGACATGAGAGGCGCTTGAATGCCGAGACCGG 103
Db 292 GCCCTGGCAGTGCAGGCGGTTCGATGACATGAGAGGCGCTTGAATGCCGAGACCGG 351
Qy 104 GTAGGGGTGCAGACTTGTCTGCTGAGAAATTTCAACGAGAGCTGCTTCATTTAG 163
Db 352 GTGGGGGTGCAGACTTGTCTGCTGAGAAATTTCAACGAGAGGCGGCTTCATTTAG 411
Qy 164 AACCTCGGCGGCGGCTTCGCGGAGAGACCTATTATACCTACATGCTGTCTGATGTC 223
Db 412 AACCTCGGCGGCGGCTTCGCGGAGAGAAATTCATACACCTACATTTGCCCGCTCTGCTC 471

```

```

OY 224 TCTGTCAATTCCTACCGAGACCTACAGATCTACAGCCGCGACATATGGAACGCTACCGT 283
DB 472 TCTGTCAATTCCTACCGAGACCTGACATCTACAGCCGCGACATATGGAACGCTACCGT 531
OY 284 GGTGTCAATTCCTACCGAGACCTACAGATCTACAGCCGCGACATATGGAACGCTACCGT 531
DB 532 GGTGTCAATTCCTACCGAGACCTACAGATCTACAGCCGCGACATATGGAACGCTACCGT 531
OY 344 GCACCTTCTGTAAGACGCTGCGAGACGAGCAGTGTATGTTCTGGAAGAGTGGGCGAGC 591
DB 532 GCACCTTCTGTAAGACGCTGCGAGACGAGCAGTGTATGTTCTGGAAGAGTGGGCGAGC 591
OY 404 AAGACAGAGGCGCACCAAGAAACACTCTCAGTTCTATGAGAGACCTGCGCCACCCCGTGA 463
DB 652 AAGACAGAGGCGCACCAAGAAACACTCTCAGTTCTATGAGAGACCTGCGCCACCCCGTGA 463
OY 464 CGGGGTGGCGGAGTGGCAGAGACGCTGTTGACAGAGACCCGCTGTTAGAGCCCTTTGG 523
DB 712 CGGGGTGGCGGAGTGGCAGAGACGCTGTTGACAGAGACCCGCTGTTAGAGCCCTTTGG 523
OY 524 AATGCCAAGACACTTCCGCAACCATTAATCTCAGCCGTTTGAAGATGATGATGATGATG 583
DB 772 AATGCCAAGACACTTCCGCAACCATTAATCTCAGCCGTTTGAAGATGATGATGATGATG 583
OY 584 TTTGACTTCAAGAGTGGCCCGCTGCGAGAGCAGATCTCAGTACCTTCTGGAAGATGTC 831
DB 832 TTTGACTTCAAGAGTGGCCCGCTGCGAGAGCAGATCTCAGTACCTTCTGGAAGATGTC 831
OY 644 CGGGGTGGCGGAGTGGCAGAGACGCTTCCAGTCTTGAAGATGATGATGATGATGATG 891
DB 892 CGAGTGGTGAACCAAGTGGGAGCGGAACTTCCACATCTTTCACAGCGCTGCGAG 951
OY 704 GGGGGCGAGAGAGACTCTCCGCTGGCTGGCTTGAAGAGGAAACCCCGACAGCTACTTG 763
DB 952 GGGGGCGAGAGAGACTCTCCGCTGGCTGGCTTGAAGAGGAAACCCCGACAGCTACTTG 763
OY 764 TACCTGTGAAGAGGCGAGTGGTGGCAAGGTCCTCCATCAGAGAGAGATGATGATGATG 1011
DB 1012 TACCTGTGAAGAGGCGAGTGGTGGCAAGGTCCTCCATCAGAGAGAGATGATGATGATG 1011
OY 824 GTTATGAGAGAGCGGCTGCTCCATGATCTCAGTGGAGAGATGATGATGATGATGATG 1071
DB 1072 GTTATGAGAGAGCGGCTGCTCCATGATCTCAGTGGAGAGATGATGATGATGATGATG 1071
OY 884 AGCATGTGGCGGAGCTCTACATCTGAGGCAACATCTCCTGCTGAGAGAGAGAGAGC 943
DB 1132 AGCATGTGGCGGAGCTCTACATCTGAGGCAACATCTCCTGCTGAGAGAGAGAGAGC 943
OY 944 AATGCCAAGACACTTCCGCAACCATTAATCTCAGCCGTTTGAAGATGATGATGATGATG 1191
DB 1192 AATGCCAAGACACTTCCGCAACCATTAATCTCAGCCGTTTGAAGATGATGATGATGATG 1191
OY 1004 GGTACACACACTTGAAGAGCCCTGAGCCAGAGAGAGATCTCCGCAAGAGAGAGAGC 1251
DB 1252 GGTACACACACTTGAAGAGCCCTGAGCCAGAGAGAGATCTCCGCAAGAGAGAGAGC 1251
OY 1064 CTGAGCCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 1311
DB 1312 CTGAGCCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 1311
OY 1124 TACAGCCGCGAGACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1371
DB 1372 TACAGCCGCGAGACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1371
OY 1184 GCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1431
DB 1432 GCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1431
OY 1244 GAAGTGTTCACATTAACAGCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1491
DB 1492 GAAGTGTTCACATTAACAGCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1491

```

```

OY 1304 CAGCAGCTCTTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1363
DB 1552 CAGCAGCTCTTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1611
OY 1364 ATCCGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1423
DB 1612 ATCCGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1671
OY 1424 AAGTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1477
DB 1672 AAGTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1727

```

## RESULT 5

ABL19551

ID ABL19551 standard; DNA; 3406 BP.

AC ABL19551;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 10126.

KW Drosophila: developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

OS Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX Claim 1; SEQ ID NO 10126; 21bp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in interactional biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (AB101840-AB101845), expressed DNA

XX sequences (AB101840-AB101845) and the encoded proteins

XX (AB101840-AB101845).

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp://wipo.int/pub/published\_pcl\_sequences.

XX Sequence 3406 BP; 885 A; 893 C; 924 G; 704 T; 0 other;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Db 176 TGGAGTCAGGACTTTGTGCTCCTCGAAGACTACGAGCGGAAGAAAGCCTTCATCGGGA 235  
Oy 166 CCTCCGGCGGCGTTCCGGAGAACTTATATACCTACATGCGTCTGTCTAGTCTC 225  
Db 236 TCTGAAGAAAGC-----ACCTACATTTGGCCAGGTGTGATCTC 273  
Oy 226 TGTCAATCCCTACCGAGACTACAGATTCACAGCCGGGACGATATGAAACGCTACCGTGG 285  
Db 274 CTTGAATCCCTACAGACACTGCCCCATCTACACCGATGACATGTCAAGGCGGTACAGAAA 333  
Oy 286 TGTCAAGTTCTATGAAGTACCACTCATTTGTTTGACAGTGGCTGACACGTATACCGGGC 345  
Db 334 CAAGCACTTCTAGAGATCCCCACACATCTTTGGGTGACGAAACGCCCTTCGGTTTC 393  
Oy 346 ACTTCCTACTGAGACGTCGGGACGAGCGATGATTTCTTGAGAGAGTGGGCGAGCAA 405  
Db 394 GCTGATCGAGAGAAACCGGGGCCAGTGCCTGCTCATCTCCGGAGAGATGTGTCGGGCAA 453  
Oy 406 GACAGAGGCCACCAAGAGACTGCTCCAGTTCTATGCAAGAGACCTGCCCAAGCCCTTAAGC 465  
Db 454 GACAGAGGCTCCCAAGAGAGTGTGCAATAGCCGCTGCTCCGGCAACCAAGACGAC 513  
Oy 466 GGGTGGCGGACGTCCGAGACCGCTGTTGCAAGCAACCCGCTTTAGAGGCTTTGGGA 525  
Db 514 CGTGAGGGCGTCAAGAGACAGCTGCTTAAAGCAATCCGTCGTGAGAGCCCTTCGGCAA 573  
Oy 526 TGCAGAGCTCTCCGCAAGATTAATCCAGCGGTTTGGAAAGTACATGATGATGAGTT 585  
Db 574 TGCAGAGCAAAACCGCAATGACAACTCTCGGCTTCGGCAAGTACATGAGACTTCCAGTT 633  
Oy 586 TGAATTCAGAGGTTGCCCGGAGGAGCCACATTCAGTTACCTTCGTAAGAAAGTCCCG 645  
Db 634 CATTTTCAAGAGAGCTCCGATCGAGGCAACATCTAAACTATCTCTGTGAAAAGTCGG 693  
Oy 646 GGTGGTGCACCAAAATCAGGAGAGAGGAACTTCCACGCTTTTACCAGCTACTGAGGG 705  
Db 694 AGTGGTGCTCAATGAGAGGAGGCAACATCTTCTCACTTCTTCACTCTTGCCCG 753  
Oy 706 GGGCGAGGAGAGACTCTCGTGGGCTTGGAGCGAAACCCGACAGAGTACTTGA 765  
Db 754 CGCGAGAGAGGCGCTTCTGAGAGAACTGGCGCTGGAGCGGCTGTGACACTTATAGCTA 813  
Oy 766 CCGTGTGAAGGGCAGTGTGCCAAGTCTCTTCATCAACAGACAGAGTGAAGGT 825  
Db 814 CCTGACAGATGGGCTCAATGGCACCTGACGAGAAATTAAGATGACGAGACTTCAAGCA 873  
Oy 826 TATGAGAGAGGCGCTCCGTCATTGATTCATCTCACTGAGATGAAAGTGAAGACTGCTCAG 885  
Db 874 GGTCCAGCAGGCTCTCAGTGTGATTCATTTCAACAGAGAGGACACGCGAGATCTTCGG 933  
Oy 886 CATCGGGCAGGCTCTCATCTGGCAACATCCACTTTGCTGTGAGAGACAGCAA 945  
Db 934 AATCGTGGCAGCATTTCTGATCTTAGGAAACGTTGGCTTCA---CTGAGTGGAGGGCAA 990  
Oy 946 TGCACGAGTTACTACTGAGAACCCAGCTCAAAATATCTGACAGGCTCCTTGTGTGAAG 1005  
Db 991 TGCAGAGTGAACAGAGGATCTGTAGTCAACCGCTGCTCGTTCTGCTGCTTAACCG 1050  
Oy 1006 TACACACTTAGGGAAGCCCTGACCAAGAGATCATGCGCAAGGGGAGAGACTCCT 1065  
Db 1051 AAGCGAATGGAAGCGCCTTAACGACCGCAAAATTAAGACGTCGCGGAGATGTGTGTAC 1110  
Oy 1066 GAGCCCACTGAACCTTGAACAGGCGGATATGCAAGGATGGCTTCCCAAGGCTGTGA 1125  
Db 1111 CTACACACTGAACAGGAGCTGCGCATCTATGCGCAGGAGCGCTTGGCCAAAGGCTGTCTA 1170  
Oy 1126 CAGCGGACATTCACCTGGCTGTGAGAAAGATCAATAGTCACTAGCGCTTAAGAGCG 1185  
Db 1171 TGAATCTGTCTCTCTGCTGTGTGAGCGCTCAACATCTCGCTCAGGCGAAGGAAAC 1230  
Oy 1186 TGAAGACCCCACTGGCGAAGCAACAGTCTTGGGCTCTCTGACATTTAGCGCTTTGA 1245  
Db 1231 GCGGGGATCC-----AGAAATATGTGATGGGCAATTTGACATCTATGCGTTCGA 1281

Oy 1246 AGTGTTCAGCATAAACAGCTTCGAGAGTTCTGCATCACTACTGCAATGAGAAGCTCA 1305  
Db 1282 GATCTTCAGAAAGAACAGCTTTGAGCAGTCTCTCATCAATTTCTGCACAGAAAGCTGCA 1341  
Oy 1306 GCAGCTCTTCATGAGCTGACTCTCAAGTCGAGCAGAGGAATACGAGCTGAGGGCAT 1365  
Db 1342 GCAGCTGTTCATGAGCTTACCTTAAATGCGAGCAGAGATGATACCGCGGAGGGCAT 1401  
Oy 1366 CGCGTGGGAACCTGTCCAGTACTTCAACAAAGATCATCTGTGACTGTGATAGAGAA 1425  
Db 1402 CGAGTGATACCTGTGAGAGTCTGCAACAAGTGATTTGCAATCTGTGACAGAGAA 1461  
Oy 1426 GTTCAAGGGCATCATCTCCATCTTGGATGAAGAGTCCGCTCTTGAGGAGCCAGGA 1485  
Db 1462 GCACAAAGGCGATATCTCCATCTACTGAGCAGAGAGTCTTGACACAGAGAGCCAGGA 1521  
Oy 1486 CTTGACCTTTCTGAGAAAGTTGAGGACACTGTCAAGCCCCACCTTCATCTTGACGA 1545  
Db 1522 TAAGACTTCTCTTGAAGAGCTCACCCAGAAATTTGGCCACACCACTACGTTTGCA 1581  
Oy 1546 CAAGCTCGTGACCAAGAACAGGAATCCCTAGACCGAGGGAGTTCCGCTTCTGCA 1605  
Db 1582 CGAAAAGGCGCCCGCCACATCAAGAAATCATGTGCGGATGAGTTCGCTTGTCGA 1641  
Oy 1606 TTAATGCTGAGAGAGTACCTACAGTGTGACTGGCTTTCTGGATAAAAAATGACCTCT 1665  
Db 1642 CTATGCCGCGAGAGTCACTACCTACAGTGTCAATGATTTCTTGACACAGACACCTGTT 1701  
Oy 1666 CTTCCGGAACCTGAAGAGAGACCATGTGACGCTCAATGAACCCCATCATGCGCCAGTCTT 1725  
Db 1702 GTTCAGAGGATCTAAAGAGAGACCTCAAGAGCTGCGCAACGGCATGTGAGAACTGCTT 1761  
Oy 1726 TGAAGAGTGAAGTCAAGTACAAAGAGGCGGAG--GAGGTTGGCCACCCAGTTCAAGAT 1784  
Db 1762 CCGGGAAGAGACTCGGACCTGGAAGCGTCCGAGACGCGTATACCCAGTTCCGGGC 1821  
Oy 1785 GAGCTCTCTGACAGCTGTGAGATCTCTGAGTCTAAGAGACCTGCTATATCCGTTGAT 1844  
Db 1822 ATCTCTTAACACTCTATGACATTTCTGATGTGCAAGAGACCCAGCTACATTTGCTGAT 1881  
Oy 1845 CAAGCCAAACGACGCCAAGCAGCGGGGTGCTTGTATGAGTGTCTCATCCGACATCAGGT 1904  
Db 1882 CAAGCCAAACGACCTGACAGCTGCCAACGCTTCAACAGATGATGTGATGACACAGGT 1941  
Oy 1905 GAAGTACCTGGGAGTATGAGAAATCTGGCGGCGCAGAGCGTGGCTTCCCTATGCTCG 1964  
Db 1942 CAAGTACCTGGGCTGATGAGAAATCTGCTGTGAGCGGAGCTGGTTGCTCTACGAGC 2001  
Oy 1965 CAATATGAGGCTTTCTGAGAGATTAACAAGTCACTGTGCCAGAGCATGAGCC----- 2018  
Db 2002 AAGCTATGAGCTTCTCTGAGAGCGCTACAAAGTCCCTGAGCAAGTCCACCTGGCCAAATTA 2061  
Oy 2019 CATGTGGGCGAGAGCGGCCAGATGCTGTGCGGCTTGGTCAAGACACTCGGCTTACAA 2078  
Db 2062 CAAGGGTCCCGGTGGCCGGAAGCGGGGTGCCAGCGCTGTGAAGAAATTTGGGCTGGGA 2121  
Oy 2079 GCCAGAAAGATCAAAATGGGCAAGACTTAAGATCTTCATTCGATCCCAAGACTTATT 2138  
Db 2122 CGAGGAAAGATCAAGGTGGGCGAGAGCAAAATTCATCTCGTGGCGGAAACCTTGTGT 2181  
Oy 2139 TGCACAGAGAGCTCCCTGGAAGTCCGGGCGGAGAGTCTTGGCCCAAGATGACGAGGGC 2198  
Db 2182 CGATACGAGAGATGCTTACAGAGAAAGAAACATGAGATGCGGCCCATCTTCCAGGCCCA 2241  
Oy 2199 CTGAGGCGCTTTCATTTGCGGACAGAAATTTCTCCGGGTGAAGCGATCAGCATCTGTAT 2258  
Db 2242 CTGAGAGGATGATGCAAAAGAGGAAGTATTTGAAGCTCGTGTCTCAGGTGATCATAT 2301  
Oy 2259 CGAGTATGTTGGCTGGGCACTAGTGGCGCGAGAGAGGCAAGAGAGAGTGGGAGC 2318  
Db 2302 GCAGACTATTTGTGCGGAAAGTTGGCAGCAGCGGGGCCAAGAGGCGAGGAGCGCC 2361

PA (PEKE) PE CORP NY,  
XX  
PI Venter JC, Adams M, Li FWD, Myers EW;  
XX  
DR WPI: 2001-656860/75.  
XX P-FSDS; ABB60812.  
XX  
PR New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions -  
XX  
PS Claim 1; SEQ ID NO 9227; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent.  
CC  
CC capable of detecting 1000 or more genes from Drosophila.  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher and in eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB101940-AB116175) and the encoded proteins  
CC sequences (AB857731-AB872072).  
CC (AB857731-AB872072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 3349 Rp. 021 + 022

Query Match	33.0%	57.0% G; 69.9% T; 0 other;
-------------	-------	----------------------------

Best Local Similarity	23.98;	Score 873.8;	DB 23;
Matches 1700; Conservative	58.18;	Pred. No. 9.8e-236;	Length 3349;

Sequence 3340 bp. 571  
aa  
SD  
Sequence 3340 bp. 571  
aa  
SD  
Sequence 3340 bp. 571  
aa  
SD

Query Match 23.0%

Best Local Similarity	23.98;	Score 873.8;	DB 23;
Matches 1700;	Conservative	0.01	
		Pred. No. 9.8e-236;	Length 3349;

CGAGTGACCATGGAGAACCCTTGA

mismatches 1172; Indels 53; Gaps 7

78 CAACTGAAATGGACAGCGGCCTGCACGAGCGGTGATCCGTCCTTCTC 124

[illegible]

138 CTCCTCGAGACTACGAGACGGAAGCCCTCATCGGGAATCTGAAGANGCC-  
184

185 GAGAACCTATTATACATACATCGCTCTGCTTACTGCTCTGTCATATCCCTACCGGAGC 344  
191 -----

245 CTACAGATCTTCCTCCTGTAATCCCTCAACGAG 235  
-----ACCTACATTGGCCAGGTCTTGATCTCTCGTGAATCCCTCAACGAG 244

236 CTGCCCATCTACACCGGATGGAACGCTACCGTGGTGTCACTTCTATGAGTA 304

[illegible]

296 CCCCACACATCTTTGGGGTACTGACACCCGCTGGTCTGACTGAGCGTGG 364

365 GACCAGCGATGATGATTTCGTGAGAGAGTGGGGACCCCACTTTT  
| |||| |  
355

356 GGGCAGTGCCTCATCTCCGAGAGAGTGTTCCGGACAAAGAGA  
|||||  
|||

424

425 CTGCTCCAGTTCTATGAGAGACTGCCAGCCCTGAACGGGGTGGCCGAGTGGGCTCCAGCAAG 415

41b GTCTGCAGTTCATAGCCGCTGCTCCGGCAACGACGACGACCGCTCAACCGG 484

[illegible]

545 GATTAACATCCAGCGCCGTTGAGGCGCTTCGGCAATGCCAAGCAACCGCAAT 535

536 GACAACCTCCGCGGCTTGGCCCAAGGCGCC 604

605 GTGGAGGCCACATTCGATTACCTCCCTGG...  
 .....GCGCAGACATGGACATCCAGTTCGATTCAAAGSAGCTCCG 595

596 ATCGGAGGCACATCTTAACATATGTGGTGAACACTTTATTTTATTTT  
664

665 GGAGAGCGCACTTCCACGCTCTTTACCGAGCTACTGACCCCCCGGCGGCGAGTGTGCGCTCAAAATGGGA 655

...GGAGGAGAGCTCTC 724

Db 656 GCGAGCGCAACTTCACACTCTTTCACACTCTTGGCCGCGCGAGCAGAGCCCTTCG 715  
 QY 725 CGTCGCTGGGCTGGAGAACCCCGAGACTTACTTACTGGTGAAGGCCAGTGT 784  
 Db 716 CAGGAAGCTGGGCTGGAGGGGCTCTGACACTTATAGTACTCTGACAGATGGGCTCAT 775  
 QY 785 GCCAAGGCTCTCCATCATCAGACAAGAGTACTGGAGGTTATGAGAGAGCGCTGTCC 844  
 Db 776 GGCACCGCTGACGAATTAACGATGGGACACTTTCACAGAGTCCAGCAGCTCTCACT 835  
 QY 845 GTCATTTGACTTACGAGATGAAGTGAAGAGACTTCTAGATGCTGGCCAGCGCTCTA 904  
 Db 836 GTGATGATTTTACCAAGAGAGAGAGCCGAGATCTTGGATGCTGGCCAGCATTTG 895  
 QY 905 CATCTGGGCAATCCACTTCTGCTGACGAGAGAGACAGCAATGGCCAGGTTTACTGAG 964  
 Db 896 CATCTAGGAAGCTTGGCTTCACTAGGTGAGG---GCAATGCCAAGGTGAACAGCAG 952  
 QY 965 AACCACTCAAAATATCTGACAGGCTCTTGGTGGAGGTACACACTTAAAGGAGCC 1024  
 Db 953 GATCTGATGATCAGCGCTCTGTTTGTGGGTGAACGCAAGCGAGTGGAGCGCC 1012  
 QY 1025 CGAGCCCAAGGAAGTCACTCCCAAGGGGGAAGACTCCTAGGCCACTGAACCTTGA 1084  
 Db 1013 CTAAAGCACCGACAAATTAACCTGGGAGATGTGTGACTTCACTCACTGAACCGAG 1072  
 QY 1085 CAGGGGCAATATGCAAGGATCGCTTGGCAAGGCTGTGTAGAGCCGACATTCACCTG 1144  
 Db 1073 CTCGCATTTATGCCAGGAGCGCTTGGCCAGGCTGTGTATGATGTCTGTTCTCTCG 1132  
 QY 1145 CTGCTCAGAAAGTCAATAGTCTACTGGCTCTAAGGACGTGAAGAGCCCACTGGCGA 1204  
 Db 1133 CTGGTTCAACGCGCTCAACATCTGCTGCGAGGAGAGAAACGCGGCACTCAGA----- 1186  
 QY 1205 AACACACGCTTCTGGGCTCTGACATTTAGGCTTTGAAGTGTGTAGATCAACAGC 1264  
 Db 1187 ---AATATGATGGGCAATCTGACATCTATGCTTGAAGTCTTCCAGAAAGACAC 1243  
 QY 1265 TTGAGACAGTTCGATCACTACTGCAATGAGAACTGCAGAGCTCTTCATCGAGCTG 1324  
 Db 1244 TTGAGACATTTCTGCATCAATTTCTGCACAGAAAGCTGCAGAGCTGTTCATCGAGCTT 1303  
 QY 1325 ACTCTCAAGTTCGAGAGAGGAATAGAGGCTGAGGGCATGCGGTGGAGACCTGTCCAG 1384  
 Db 1304 ACCCTTAAGTCGAGAGAGATGAATACCGCGGAGGCGCATGAGTGAATACCTGTGAG 1363  
 QY 1385 TACTTCAAAAGAATCTCTGTGACCTGTGTAGAGAAAGTTCAAGGCGCATCTCC 1444  
 Db 1364 TACTTCAAAAGAATGATTTGCAATCTGTGAGAGAAAGCAAGGCGCATCTCC 1423  
 QY 1445 ATCTTGAAGAGAGTGCCTGCTGGGAGAGCCAGGAGCTTTCGGAGAG 1504  
 Db 1424 ATACTGAGAGAGAGTGTCTGCGACAGAGAGAGCCAGGATTAAGACTTCTCTTGAGAA 1483  
 QY 1505 TTGGAGACACTGTCAAGCCCAACCTCTCTGAGCGACAAAGTCCGTCACCGAAG 1564  
 Db 1484 CTCACCCAAAGATTTGGCCAGACACCACTAGCTTGGCACAGAAAGCGCCGCCAC 1543  
 QY 1565 ACCAGGAATCCCTAGACCGAGGAGGTTCCGCTTCTGCATTTACTGTGAAGGTGAC 1624  
 Db 1544 ATCAAGAAATATATGCTGCGATGAGTTCGCTTGGTCACTATACCGCGACCTCAC 1603  
 QY 1625 TACAGGTGACTGGTCTTGGATAAACAATGACCTCTCTTCGGGAACCTGAAGAG 1684  
 Db 1604 TACAGGTGATGATGATCTCTGACAAAGAACAGACCTGTGTTCAGGAGATCTAAAGAG 1663  
 QY 1685 ACCATGTGACGTCAATGAAGCCCATCATGAGCCAGTCTTGAACAAGATGAGTCACT 1744  
 Db 1664 ACCCTGAGCAAGGCTGGCAAGGCAATTTGTAGGAAGTCTCCCGGAAGAAGAGTGGCG 1723  
 QY 1745 GACAAAGAACGGCCAG-GACGGTGGCCACCCAGTTAAAGATGAGCTCTCTGACAGTCTG 1803  
 Db 1724 ACCCTGAAGCGTCCGAGAGCGCTATCACCAAGTTCCGGGCACTCTCTTAACAACCTCATG 1783  
 QY 1804 GAGATCTGAGGTCTAAGAGACCTGCTATATCCGGTGCATCAAGCCAAACGAGCCAG 1863  
 Db 1784 GACATCTGATGTCGAAGAGCGGAGCTACATTCGCTGATCAAGCCAAACGAGCTGAC 1843  
 QY 1864 CAGCGGGTCCGCTTTATAGAGTGTCTATTCGACATCAGGTGAAGTACCTGGAGTGA 1923  
 Db 1844 ACTGCCAAGCTCTTCAACAGATGATGTTGATCTGACACAGGTCAAGTACCTTGGCTGAT 1903  
 QY 1924 GAGATCTGCGGTGGCGAGAGCTGCTTGCCTATCTGTCGAAATATGAGCTTTCCTG 1983  
 Db 1904 GAGAAATCTGCTGTGAGCGAGCTGTTCGCTTACCGAGAAACGTAAGAGCTTCTG 1963  
 QY 1984 CAGAGGTAAAGTCACTGTGCCAGAGACATGAGC-----CATGTGGCGAGCGGCC 2037  
 Db 1964 GAGCGTACAAAGTCCCTGAGCAAGTCCACCTGCGCCCAATTTACAGAGGTCCCGGTGG 2023  
 QY 2038 CAGATGTTGTGGCGCTGTTGTCAAGACCTCGGCTACAGGCGAAGAGTACAAATG 2097  
 Db 2024 AAAGCGGTGTCCAGCAGCTGGTGAAGATTTGGCTGGGACGAGGAAAGTACAGGCTG 2083  
 QY 2098 GCGAGACTAAGATCTTCAATCCGATTCGCCAAGACCTTATTTGCCAGAGACTCCTG 2157  
 Db 2084 GCGGAGACGAACTATTCATCCGCTGGCGAGAACTTGTTCATCGAGAGATGCTTAC 2143  
 QY 2158 GAAGTCCGCGGAGAGTGTAGCCACCAAGATCCAGCGGCTGAGAGGGCTTTCATTTG 2217  
 Db 2144 CAGGAGAAAGAAATGATGATAGCGGCGCATTCACAGGCCACCTGGAAGGATTTGATGCA 2203  
 QY 2218 CGACAGAAATTTCTCCGGTGAAGCGATCAGCATCTGATTCAGTCAATGAGTGGCGT 2277  
 Db 2204 AGGAGGAGATATCTGAACACTGCTGCTCAGGTGATCATATGAGAGCTATTTGCGCGA 2263  
 QY 2278 ACATGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2337  
 Db 2264 AAGTTGGACAGAGGCGGCGCAAGAGCGAGGAGGCGCGCATTAAGATCTGCTCTT 2323  
 QY 2338 ATCCGTCCTTCAATTTGGCCATTCACCCGCTGAGTGAAGTCTTCTTTGAG 2397  
 Db 2324 ATCAAGGGCTTCAATCCCGGAACGATGGCCGAAATGCTTCAATGAAGGATTTATGGC 2383  
 QY 2398 CAGTCCGCGCTCATATTTTGTAACTGAGGCGCAACTGCCCGGAATGTTCTGAC 2457  
 Db 2384 AAGCGCAAGCGTATGTGGCTGTTCGATGGCGAAGAGGCTGCCCAAGAGTACTGAT 2443  
 QY 2458 ACTCTGCGCCACACCCCACTGCTGAGAGAGGCTCAGAACTGCTACGGGAATG 2517  
 Db 2444 AAGAGTGGCCCATGCTCCCTGGGCACTGCGAGAGGCGCTCCGCAATCTTCAACGCTG 2503  
 QY 2518 TGCATGAAGAACTATGAGTGAAGTACTGCGGAGCATCAGCCCTGAGTGAAGAGCAG 2577  
 Db 2504 CACCTCTCCTTCTGCGCAGAGATCTACCGCTTAACTGACGCGCGAGCAAGAAAGAG 2563  
 QY 2578 CTGCAGCAAAAGGCGGTGCTAGTGAATTTTCAAGGCGAAGAGCAACTACCCCGAG 2637  
 Db 2564 TTGAGACTGAAGGTCTTGGGAGAGAGTCTTAAAGGCAAGAAAGCAACTACGAGAG 2623  
 QY 2638 ACTGTCCCAAGACTTTCATTAACACAGGCTTGGCAGAGAGAGA-----TCAGC 2688  
 Db 2624 AGTGTCCACTTGTGTTCCAGAGGAGCCGATCCCAAGAGGACATCCAGCGAGTCAAC 2683  
 QY 2689 CCCAGAGTCTTCAATCTTGGCTGTAACCCATCCAGTACCTGCTGCGCGGTATAAA 2748  
 Db 2684 GACTTGTGTGGCCAGACCTTTCGCGAGAGGAGAGCTTAAGTACAGATCTTCTGACCAAG 2743  
 QY 2749 TACGACCTGAAGGTTTACAAGCCTGCGCCCGGAGCTGCTCTCAAGCCAGTGTGTG 2808  
 Db 2744 TTGATGTGGCAGCGGATCAAGTCCCGAGACCGCTTATCTCTGTGAGCAACAAGGCTATA 2803  
 QY 2809 GTCATTTGTGAGAGTCTAAAGTCAAGCAGAGATTTGATATGCG---CAACCTAACCGGA 2865  
 Db 2804 TATGTCTCTGAGCGCAAGACTTAAAGCAGAGACACCGGCTGCGGTAGACAGATCGAC 2863





QY 1515 CTGTCAAGCCCCACCTTCATCTTGTGACGACAAAGCTGCTGACCAAGAAGCAAGAAAT 1574  
 Db 3486 AATTGGCCACACACACACACTACGTTTGCACGAAAGCGCCGCCACATCAAGAAAA 3545  
 QY 1575 CCCTAAGACCGAGGGAGCTTCCGCCCTTGTGCATATATGTGGAGAGGTGACCTACAGTGTGA 1634  
 Db 3546 TCATGCTGCGGAGTAGTGTGCTGTGGTGAAGTCCGGGAGGTCACTACACATGTCA 3605  
 QY 1635 CTGGGTTTCTGATAAAAACAATGACCTCTCTCCGGAACCTGAAGAGACCATGTCA 1694  
 Db 3606 ATGATTCCTGTGACAAAGAACACGACCTGTGTGTACAGGATCTAAGAGACCTCAGACA 3665  
 QY 1695 GCTCAATGACCCCAATCATGTGCCCCAGTCTTTGACAGAGTGAAGTCACTGACAGAAAGC 1754  
 Db 3666 AGGCTGGCAACGGCATTTGTGAGAACTCTTCCGAGAGAAAGAGTGGCGACCTGAAGC 3725  
 QY 1755 GGGCAG-GAGGCTGGCCACCCAGTTTCAGATGAGGCTCTGACGTCTGACGATCTCGA 1813  
 Db 3726 GTCCGGAGACGGCTATACCCACAGTTCGGGCACTCTTAACAACCTCATGACATTTCTGA 3785  
 QY 1814 GGTCATAAGGACCTGCTCCATATCCGGTGCATCAAGCCAAAGACGCCAAGACCGGGTTC 1873  
 Db 3786 TGTGCAAGGACCCGACGACTACATTTCCCTGCATCAAGCCAAAGACCTGACACTGCAACG 3845  
 QY 1874 GCTTTGATGAGGTGCTCATCCGACATCAGTGAAGTACCTGGGACTGATGAGAAATCTGC 1933  
 Db 3846 TCTTCAACGATGATGTGTGACTGACACAGTCAAGTACCTTGCCCTGATGAGAAATCTGC 3905  
 QY 1934 GCGTGGCAGAGCTGCTGTTGCTTATGCTGCAATATGAGCTTTCTGACAGAGTACA 1993  
 Db 3906 GTGTGAGCGAGCTGTGTTCCCTACCGACGAAACGATGACCTCTGCGGCCCTACA 3965  
 QY 1994 AGTCACTGTGCCAGAGACATGAGCC-----CATGTGGCGAGAGGCCCCAGATGTGAG 2047  
 Db 3966 AGTCCCTGAGCAAGTCCACCTGCGCCCAATTACAGGGTCCCGGTGGCCGAAAGCGGTG 4025  
 QY 2048 TGGCCGTGTTGGTACAGACCTCGGCTACAAAGCCAGAAAGTACAAATGGGCGAGACTA 2107  
 Db 4026 TCCAGACAGCTGTGAAGATTTGGGCTGGGAGAGGAAAGTACAGGGTGGCGGAGACA 4085  
 QY 2108 AGATCTTATCTCGATTTCCCAAGACCTTATTTGCCACAGAGAGACTCCCTGGAATCCGCG 2167  
 Db 4086 AACTATTCTATCCGTCGGCGGAGAACCTTGTTCGATACGAGATGCTTACCAAGGAGAA 4145  
 QY 2168 GGCAGAGTCTAGCCACCAAGATCCAGCGGGGCTGGAGGGCTTTCATTTGGGACAGAAAT 2227  
 Db 4146 AACATGAGATAGCGGGCCATCATCCAGGCCCACTGSAAGGATTTGATGCAAAAGAGAAAT 4205  
 QY 2228 TTCTCCGGGTGAAGGATCAGCCATCTGTATCCAGTCAATGCTGGCGGCACTGGGCC 2287  
 Db 4206 ATCTGAACCTGCTGCTCAGGTGATCATATGACAGAGCTATTTGTGCCGAAAGTTGGCAC 4265  
 QY 2288 GGAGAGAGGCGCAAGAGAGAGTGGGACGCCAGACCATCCGTCGACTCATCCGTGCT 2347  
 Db 4266 AGCAGGGCGGCAAGAGAGGAGGAGGCCGATGAATTCGCTTATCAAGGGCT 4325  
 QY 2348 TCATTTTGGCCATTCACCCCGGTCCTCGAATGCTCTTCTTGTGGACCAAGCGGCGG 2407  
 Db 4326 TCATCACCAGGAGAGATGATCCCGAATGCTTCATGAAGAGTTTATTTCCCAAGCCCAAGC 4385  
 QY 2408 CCTCATTTTGTCTTAACCTGAGAGGCGGCACTGCCCCGGAATGTTTGTGACACCTTCGCG 2467  
 Db 4386 GTATGTGCTGCTTGCATGTGCGAAGAGTGTGCCACCAAGTACTGATTAAGACTGGC 4445  
 QY 2468 CCACACCCCACTGCTGAGAGAGGCTCAGAACTCTTACGGGAACTGTGCTGATGAAGA 2527  
 Db 4446 CCGATGCCCCCGGACGACGCGAGAGGCTCCGGCATCTTTCACCGCTGACCGCTCTCC 4505  
 QY 2528 ACATGCTGTGAAGTACTGCGGAGACATCAGCCCTGAGTGGAGCAGAGCTGACGCAAA 2587  
 Db 4506 ACTTGCGCAGATCTACCGCTTAACTGACGCCGACAGAGAGGCAATTCGAGCTGA 4565

QY 2588 AGCGGTGCTAGTGAATTTTCAAGGCGAAGAGACACTACCCCAAGTGTCCCA 2647  
 Db 4566 AGGTCTGGCGGAGAGGTCTTTAAGGCGAAGAAACAATACGACGAGCACTGTGCA 4625  
 QY 2648 GACTTCTATTAGCACAGCGCTTGGCACAGAGAGA-----TCAGCCCCAGAGTCC 2698  
 Db 4626 CTGGTTCAGAGAGGAGCGCATCCCAAGAGACATCCAGGAGTCAACGACTTGTGG 4685  
 QY 2699 TTCAATTCCTGGGCTGTGAACCCATCCAGTATGCCGTGCTGTAATAATACGACGTA 2758  
 Db 4686 CCAGACCTTGGCAGCGCAGCAGCTTTAAGTACCATCTCTTGACACCAAGTTCCATGCGC 4745  
 QY 2759 AGGTTACAAGCTTCCGCCCGGCGAGCTGCTCAGCCGCCAGCTGTGTGATTTGG 2818  
 Db 4746 ACGGTACAAAGTCCCGCACCTTTATCTGCTGACGAAACAAGCTATATATGTCCTCG 4805  
 QY 2819 AGGATGCTAAAGTCAAGCAGAAATTTGATTC---CACTTAACCGGAATCTCTGCA 2875  
 Db 4806 ACGCAAGACCTACAAAGCAGAGACACCGCTGCGCTGACAAAGATCGACTTCAGCTGA 4865  
 QY 2876 GTAGCCTGAGTGAAGCTTATTTGTGCTTACAGTGCAGCGTGAAGACAACAAGAGAG 2935  
 Db 4866 CGAACCAAGACAGACCGATGATGTCATCCGATACCGCTGACCTGAATAAAGACAAAG 4925  
 QY 2936 GAGATGTGCTGCTGCAAGTGAATCATGTGATCGAG 2970  
 Db 4926 GCGACTGATTTCTGATTCATTCGCGCATATATCGAG 4960

RESULT 8  
 ABL04914  
 ID ABL04914 standard; cDNA; 7057 BP.  
 XX  
 AC ABL04914;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 9224.  
 XX  
 KW Drosophila: developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN MO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 XX 23-MAR-2001; 2001WO-US09231.  
 PF  
 XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI: 2001-656860/75.  
 DR P-PSDB: ABB60811.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 9224; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (AB85737-AB872072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published-pct-sequences.  
 XX Sequence 7057 BP: 1940 A; 1678 C; 1698 G; 1741 T; 0 other:

Query Match Best Local Similarity 20.5%; Score 749.4; DB 23; Length 7057;  
 Matches 1561; Conservative 0; Mismatches 1081; Indels 93; Gaps 7;

```

QY 317 TTTCAGTGTCTGACACTGTATACCGGGCACTTCTACTGACGCTCGGACAGCAAGT
DB 2812 TTTCAGTGTCTGACACTGTATACCGGGCACTTCTACTGACGCTCGGACAGCAAGT
QY 377 ATGATTTCTGGAGAGAGTGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
DB 2872 CTGATCTCCGAGAGAGTGGTTCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
QY 437 TATGAGAGAGTCTGCCAGAGCCCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
DB 2932 ATACCGCCGCTGCTCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
QY 497 AGCAACCCGCTGTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
DB 2992 AGCAATCCGCTGTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
QY 557 CGGTTTGAGAGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
DB 3052 CGGTTTGAGAGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
QY 617 ATTCAGAGTTTACCTCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
DB 3112 ATCTTAAACATATCTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
QY 677 TTCCAGCTCTTTTACCAAGTACTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
DB 3172 TTCCAGCTCTTTTACCAAGTACTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
QY 737 TTGGAACGGAACCCCAAGTACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
DB 3232 CTGGAGCGGGCTGAGACCTTATAGCTACCTGAGAGGAGGAGGAGGAGGAGGAGG
QY 791 -----
DB 3292 GGGTATTGACAAATTGGTCTGTGATCTTGAAGTTCTAATCCAGAGCTCAATGGCACCCTGA
QY 795 CCTCATCAGCAGCAGAGAGTACTGAGAGGTTATGAGAGAGGCGCTGTCCGCTATTGACT
DB 3352 CGAGAAATTACAGATGCGGAGCAGCTTCAAGCAGGTCAGGAGGCTTCACTGATCGAAT
QY 855 TCACCTGAGATGAGTGGAGGAGTGTCTGAGATCTGTGCGAGGCTCTACATTTGGGCA
DB 3412 TCACCAAGGAGAGCAGCGGAGATGCTGGAATCGTGCCAGCAATCTGTGATCTAGGAA
QY 915 ACATCCACTTTGCTGAGAGGAGAGCAGCAATGCCAGGTTACTAGTGAAGAACCACTCA
DB 3472 ACGTTGGCTT---CACTGAGGTGGAGGCAATGCCAGGTTACTAGTGAAGAACCACTCA
QY 975 AATATCTGACAGAGCTCTTGTGTGAAAGTTACAACTATTAGGAGAACCCGACCCACA
DB 3529 TCACCCGCTGCTGTGTGGGTGTAAACGCAAGCAGCAAGTGAAGCCGCTTAAAGCCAC
QY 1035 GGAAGATCTATCCCAAGGGGAGAGAGCTCTGAGAGCCCACTGAGAACCTTAAACAGC
DB 3589 GCACAAATTGAGCTGCGGAGAGTGTGTGACCTCAACCACTGAGAGAGAGCTTCCCAT
QY 1095 ATGCAAGGAGTGGCTTCCCAAGGCTGTGTACAGCCGAGCACTTACCTGGCTGTGAGAA
DB 3649 ATGCAAGGAGAGCTTGGCCAGGCTGTGTATGATCGTGTCTCTGTGCTGTGCTGAC
QY 1155 AGATCATTTAGTACGTGGCTCTAAGAGCCTGAGAGCCCGACCTGCGCAAGCAGCAGG
  1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111

```

```

DB 3709 GCCCTAACATCTCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
QY 1215 TTCTGGGCTCTCTGACATTTTACGCTTTTGAAGTCTTTGACATTAACAGCTTGA
DB 3760 TGATGGGCTATTTCTGAGACATCTATGAGGCTTCTGAGATCTTCTGAGAGGAGG
QY 1275 TCTGCTATCACTATCTGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
DB 3820 TCTGATCAATTTCTGCAACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
QY 1335 CGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
DB 3880 CGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
QY 1395 ACAAGATCATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
DB 3940 ACAAGATCATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
QY 1455 AAGAGTCTGCTGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
DB 4000 AAGAGTCTGCTGCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
QY 1515 CTGTCAAGCCCGACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
DB 4060 AATTGGCCGAGCAGCAGCAGTACCTTTGAGGAGGAGGAGGAGGAGGAGGAGGAGG
QY 1575 CCTAGACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
DB 4120 TCATCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
QY 1635 CTGGGTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
DB 4180 ATGATTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
QY 1695 GCTCAATGAACCCCAATCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
DB 4240 AGGCTGGAACCGCATTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
QY 1755 GGGCAG-GAGGCTGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
DB 4300 GTCCGAGAGCGGCTATATCAACCAAGTTCGGGCAATCTTAACAAACCTTATGACAT
QY 1814 GGTCTTAAGAGCCCTGCTATATCCGCTGATCAACCAAGGAGGAGGAGGAGGAGGAGG
DB 4360 TGTCAGAGAGCCGAGCTATATTCGCTGATCAACCAAGGAGGAGGAGGAGGAGGAGG
QY 1874 GCTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
DB 4420 TCTTCAACGATGATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
QY 1934 GCGTGGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
DB 4480 GTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
QY 1994 AGTCAGTGTGGCCAGAGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
DB 4540 AGTCAGTGTGGCCAGAGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
QY 2048 TGGCGGTGTGTGAGAGCAGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
DB 4600 TCAAGCAGTGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
QY 2108 AGATCTTCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
DB 4660 AACTATTATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
QY 2168 GGCAGAGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
DB 4720 AACATGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
QY 2228 TTCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
DB 4780 ATCTGAACGTGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
  1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111

```

QY 2288 GGAGGAGGCGACCGAAGAGAGTGGGACGCCAGACCATCCGTCGATCCGTCGCT 2347  
 DB 4840 ACCAGCGCGCCGAAGAGGCGCAGAGGAGCGCGCATTAAGATCGTCTTATCAAGGCT 4899  
 QY 2348 TCATTTTGGCCATTCACCGCGTGGCCGAGATGCCCTTCTTGGACCAAGTCGCGG 2407  
 DB 4900 TCATACCCCGAGACATGCGCCGGAATGCTTCAATGAAGATTTATTTGCAACGCCAAC 4959  
 QY 2408 CCTCATTTTGGCTTAACCTGAGAGCGGCACTGCGGAAATTTCTGACACCTCCGTCG 2467  
 DB 4960 GATATGGCTGCTTGCATCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5019  
 QY 2468 CCACACCCCGACCTCCCTGAGAGAGGCGCTGCAAGCTGCTGAGGAGGAGGAGGAGGAG 2527  
 DB 5020 CCGATCGCCCTGGGACACTGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5079  
 QY 2528 ACATGCTGAGAGATGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2587  
 DB 5080 ACTTGGCCAGGATCTACCGCTTAAACTGACCGCGAGCAGAGAGGAGGAGGAGGAGGAG 5139  
 QY 2588 AGCGGTGCTGATGAATTTTCAAGGCGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2647  
 DB 5140 AGCTCTGCGGAGAGAGGCTCTTAAAGGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5199  
 QY 2648 GACTCTTCAATTAGACACGCGCTTGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2698  
 DB 5200 CTGTTGTCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5259  
 QY 2699 TTCAATCTTGGGCTTGAACCCATGACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2758  
 DB 5260 CCAGACCTCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5319  
 QY 2759 AGGTTTACAGGCTGCGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2818  
 DB 5320 AGGCTACAGAGTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5379  
 QY 2819 AGGATGCTAAAGCTCAAGCAGAGAGATTTGATTTATGCTTCAACCTTACCGGAGATCTG 2875  
 DB 5380 ACGGAGAGGAGCTCAAGCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5439  
 QY 2876 GTAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2935  
 DB 5440 CGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5499  
 QY 2936 GAGATGCTGCTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2970  
 DB 5500 GCGACTTGATTTCTGATCATTTCCGCGCATTAATGAG 5534  
 RESULT 9  
 AAH02967  
 ID AAH02967 standard; DNA; 2560 BP.  
 XX  
 AC AAH02967;  
 XX  
 DT 15-JUN-2001 (first entry)  
 XX  
 DE Human shear stress-response coding sequence SEQ ID NO: 170.  
 XX  
 KM Human; shear stress-response protein; vascular disease;  
 KM arteriosclerosis; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200125427-A1.  
 PD 12-APR-2001.  
 XX  
 PF 02-OCT-2000; 2000MO-JP06840.  
 XX  
 PR 01-OCT-1999; 99JP-0280976.  
 XX

PA (KIOM) KIOMA HAKKO KOGYO KK.  
 PA (NOJ1/) NOJIMA H.  
 XX  
 PI Nojima H, Yoshisue H, Obayashi M, Ota T, Kawabata A, Sakurada K,  
 DR Kuga T, Sekine S, Nakamura Y, Sugano S;  
 DR WPI: 2001-266308/27.  
 DR P-PSDB: AAB90827.  
 XX  
 PT DNA sequences, proteins encoded by them and antibodies against them  
 PT useful in diagnosis and treatment of vascular disease caused by  
 PT arteriosclerosis -  
 XX  
 PS Claim 1; page 662-664; 678bp; Japanese.  
 CC  
 CC The present invention provides the protein and coding sequences of a  
 CC number of human shear stress response proteins. These are useful in the  
 CC diagnosis, treatment and screening of vascular diseases caused by  
 CC arteriosclerosis, including heart failure, post-PTCA restenosis and  
 CC hypertension.  
 XX  
 SO Sequence 2560 BP; 543 A; 761 C; 746 G; 510 T; 0 other;  
 Query Match 20.4%; Score 745.2; DB 22; Length 2560;  
 Best Local Similarity 88.3%; Pred. No. 1.8e-199;  
 Matches 832; Conservative 0; Mismatches 108; Indels 2; Gaps 2;  
 QY 44 GCGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 103  
 DB 277 GCGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 336  
 QY 104 GTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 163  
 DB 337 GTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 396  
 QY 164 AACCTCGGCGGCGGCTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 223  
 DB 397 AACCTCGGCGGCGGCTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 456  
 QY 224 TCTGTCAATCCCTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 283  
 DB 457 TCTGTCAATCCCTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 516  
 QY 284 GGTGTCAATCCCTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 343  
 DB 517 GGTGTCAATCCCTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 576  
 QY 344 GCACTTCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 403  
 DB 577 GCACTTCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 636  
 QY 404 AAGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 463  
 DB 637 AAGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 696  
 QY 464 CGGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 523  
 DB 697 CGGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 756  
 QY 524 AATGCCAAGACTCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 583  
 DB 757 AATGCCAAGACTCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 816  
 QY 584 TTTGACTTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 643  
 DB 817 TTTGACTTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 876  
 QY 644 CGGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 703  
 DB 877 CGGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 935  
 QY 704 GGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 763

Db 936 GGGGGGAGAGAGACTCTTTCGAGGCTGGGCTTGGAGGAGAACCCGAGAGCTATCTG 995  
QY 764 TACCTGTGAAGGCGCCACTGTGCGCAAGGTCTCTTCATCAACGACAGAGTGTGACGAG 823  
Db 996 TACCTGTGGA- GCGCAATGTGCGCAAGGTCTCTTCATCAACGACAGAGTGTGACGAG 1054  
QY 824 GTTATGAGGAAGGCGCTGTCCGTCATTTGACTTCAGTGAATGAGAGACTTGTCTC 883  
Db 1055 GTCTGAGGAGGCTCTGACAGTCATTTGATTTCACGAGATGAAAGTGGAGACCTGCTG 1114  
QY 884 AGCATCTGTGCGGCTCTCTACATCTTGGGCGCAACATCCACTTGTCTGAGGAGACAG 943  
Db 1115 AGCATCTGTGCGGCGCTCTCTACATCTTGGGCGCAACATCCACTTGTCTGAGGAGACAG 1174  
QY 944 AATGCCAGGTCTTACTGTGAACACAGCTCAATATCTGACC 985  
Db 1175 AATGCCAGGTCTTACTGTGAACACAGCTCAATATCTGAGC 1216  
RESULT 10  
ABL28593  
ID ABL28593 standard: DNA: 3874 BP.  
XX ABL28593;  
AC ABL28593;  
XX 26-MAR-2002 (first entry)  
DT 26-MAR-2002 (first entry)  
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 37252.  
XX Drosophila: developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX Drosophila melanogaster.  
OS WO200171042-A2.  
XX 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US09231.  
XX 23-MAR-2000; 2000US-191637P.  
XX 11-JUL-2000; 2000US-0614150.  
XX (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
DR WPI: 2001-656860/75.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions -  
PS Claim 1: SEQ ID NO 37252; 21bp + Sequence Listing: English.  
XX The invention relates to an isolated nucleic acid detection reagent.  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL1840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp://ipr.int/pub/published\_pcr\_sequences.  
XX Sequence 3874 BP: 1073 A; 982 C; 955 G; 864 T; 0 other;

## Query Match

Best Local Similarity 13.9%; Score 510.6; DB 23; Length 3874;  
Matches 1168; Conservative 0; Mismatches 884; Indels 37; Gaps 9;

QY 78 AGAGCGCTTGACTGGCCGAGACCGGGGTAGGGGTGACGACTTTGTCTGTGAGAAAT 137  
Db 182 AGGACACGATGCTATGCAACGGGAAGCTGGAGTGTGACGAGTGTCTCTGACCAAG 241  
QY 138 TCACAGTAGAGCTGCTTCTTATGAAACCTCCGGCGCGGTTCGGGAGAACCTATT 197  
Db 242 TGTTCATGAG--AAGTTCATGATTAACCTGAGAAACGATTCGAAACGAGATCT 298  
QY 198 ATACATCAATGCGTCTCTCTAGTCTGTCAATCCCTACCGAGACTACATCTACATA 257  
Db 299 ACATATCAATGCGGCGCAAGGTGCTGTGATGATGATGATGATGATGATGATGATGAT 358  
QY 258 GCGGCGCATATGGAAGCTACCGTGTGTGCACTTCTTGAAGTACCACTCATTTGT 317  
Db 359 GTCCGGAACCATTCGCAAAATGCAAGGTGTGGAGCTCTTTCGAGAAATGCGCACATTTGT 418  
QY 318 TTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 478  
Db 419 TCGCATTTGCGGATTTCCGCTTACAGGGGTGCTCAAGCAGCGCCACAGACCTCATTC 478  
QY 378 TGATTTTGGAGAGAGTGGGGCAGGCAAGACAGAGGCGCAAGAGACTGCTCAATTC 437  
Db 479 TAAATTCGGGTGAGTCCGCTGCGGCGCAAAACGAGGCTTCCAGATCATGAAATGACA 538  
QY 438 ATGC-----AGAGACCTGGCCAGCCCTTCAACGGGTGGCGAGTGGAGACCG-CTGT 491  
Db 539 TTGGCGGCGGTGAGAGATTCGCGAGGCTGACGAAATGCAAGAGTCAAGAACCTTTGA 598  
QY 492 TGCAGAGCAACCCCGTTTGAAGGCTTTGGGAATGCCAAGCTCCGCAAGATTA 551  
Db 599 TCCAGAGCAATGCAATTCGGAACCTTCGGAACGCAAGACTCAACGAGACACT 658  
QY 552 CCAGCCGGTTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 611  
Db 659 CCAGCGCTTCCGAAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 718  
QY 612 GCCACATTCAGTACTCTGTAAGAAATCCGGGTGTCGACAAATTCACGAGAGAC 671  
Db 719 GCATCATCAACCACTATCTGCTCGAAGTGGCGGTGTCGAGCAGCAGCAGCAGCAGC 778  
QY 672 GGAACCTTCAGTACTCTTACAGTACTGAGGGGGCCAGAGAGACTCTCCGTGGC 731  
Db 779 GGAACCTTCAGTACTCTTATCAGCTGCTGTCGACCAACAGAGCTGCGGCAAT 838  
QY 732 TGGGCTTGAAGGAGAACCCGCAAGCTTGTACCTGTGAGAGGCGCAGTGTGCAAG 791  
Db 839 ATGAACTGCAAGAGGAACGCGCAAGTATCATTGTAACCAAG-----CAGCA 889  
QY 792 TCTCTTCATCAACGACAGAGAGTGAAGTATGAGAGAGCGCTGCCATATTG 851  
Db 890 TGACATATGAGGAGAGAGTCCGACTACAGAGGCACTGCAATGCCCTGAAGACCTTG 949  
QY 852 ACTTCACTGAGATGAGAGAGTGTGAGACTGTGAGATGTCGAGGCTGCTCAATCTG 911  
Db 950 GCTTTCACGAGATGAGGTCGACGACATTTGGCCACAATTCGCGTGTTCATCTG 1009  
QY 912 GCAACATCCATCTTCTCTGACGAGCAGCAATGCCAGTTACTGAGAACCAAC 971  
Db 1010 GAAATGTGAGTTCCAACCATTTAGGA---TGAGCTGTGATCACAATAAGAGCACT 1066  
QY 972 TCAATATCTGAGCAGGCTCTTGTGTGAGAGTACACACTTGAAGGCGCTGAGACC 1031  
Db 1067 TGAATTCACGCGCCAGACTGTCTCCAGTCCGAAACCCAGAGTGTCCAGGCTGACCA 1126  
QY 1032 ACAGAGATCATCGCCCAAGGGGAGAGCTCTGTGAGCCCACTTGAACCTTGAAGGCGG 1091  
Db 1127 AACGGTCAATTCGCGCGGTGGAATGTGATGAGAGATATCAACGCGCAGGCGAG 1186  
QY 1092 CATATGCAAGAGATGCGCTTGCACAAAGCTGTGTACAGCGCATTCACCTGCGGTG 1151  
Db 1187 AGTACGGAAGATGCTCTGCGCAAGGCCATCTACAGACGCTCTTACCGATCATCT 1246  
QY 1152 GAAAGATCAATAGGTCACTGGGCTCTAAGGAGCGTGAAGGCCCAAGCTGGCGAAGACCA 1211

Db 1247 CCCGATCAACAGGGCCAAAT--CCTATTCGGTGGCAGCAAGCGAG--GCGCATTTAACT 1303  
Oy 1212 CGGTTCTTGGGGTTCGACATTTACGGCTTGAAAGTGTTCAGATACACCTTCGACC 1271  
Db 1304 CGGTATCGGAGTGTGTGATATTTACGGCTTCGAGATCTTCACAGCCACACCTTTGAAAC 1363  
Oy 1272 AGTTCTGCATCAACTACTGCAATGAGAAAGCTCAGCAGCTTCTTCAATCGAGCTACTCA 1331  
Db 1364 AGTTCTGCATCAACTACTGCAATGAGAAAGCTCAGCAGCTTCTTCAATCGAGCTACTCA 1423  
Oy 1332 ACTCGGAGAGGAGGAATACGAGGCTGAGGGCATCGCTGGGAACCTGTCCAGTACTTCA 1391  
Db 1424 AACAGGAGCAGAGGAGTACCGCTGAGGGCATGAAATGAGCAATATTGAATCTTCA 1483  
Oy 1392 ACAAGAAGTATCTGTGCTCGGTAGAGGAGATTCACAGGATTCATCTCATCTTGG 1451  
Db 1484 ACAAGAAGTATCTGTGCTCGGTAGAGGATTCACAGGATTCATCTCATCTTGG 1543  
Oy 1452 ATGAAGAGTGCCTGCGTCTCGGGAGGCGCCAGACCTGACCTTTCTGGAAGTTGGAGG 1511  
Db 1544 AGGAGGCGCTGCTCAAGCGTGGGCAAGGTAGCGAGACACATCTTGGGGGCGCATGGACA 1603  
Oy 1512 ACAGTGTCAAGCCCAACCTCTACTTCTGAGCCACAGACTCGCTGACCAAGAACAGCA 1571  
Db 1604 AAAATCTGAGAGCATCCGCACTACACCACTGCTCAGCT---GAAGCCACAGATTAAG 1660  
Oy 1572 AATCCCTAGACGAGGAGGAGTTCGCGCTTCTGCTATTCGTAGAGAGTACCTACAGTG 1631  
Db 1661 AGCTAAGCATGCGGAGGATTTCCGATACCCACTACCGCGCGATGATGATCTACACA 1720  
Oy 1632 TGAGTGGGTTTCTGATATAAAACAATGACCTCTCTCCGGAACCTGAGAGGACCATGT 1691  
Db 1721 TCAATGATTCATTCGAAAGAAAGAAAGATACGCTGATACAGACTTTAAGCGCTGCTGC 1780  
Oy 1692 GCAGTGCATGAACCCCATCATAGGCCAGTGTGTTGACAAGAGTGAAGTCACTGACACA 1751  
Db 1781 ACAACTCTAAGATCCCAACCTGAGCAGAGATGTGGCCCAAGAGACAGGATATCAAGA 1840  
Oy 1752 A-----GCGGCGAGGAGGTGGGCCACCACTGCAATGAAGAGCCCTCGAGCCCTG 1801  
Db 1841 AACCACCAAGAGACCATTTGACAGCGCGACCTTTGTCAGAGATGATGCTGATCTGG 1900  
Oy 1802 TGGAGATCCTGAGGCTCTAAGAGAGCTGCTATATCCGGTGCATCAAGCCAAAGACGCCA 1861  
Db 1901 TGGTTACTGTCTGTAAGAAAGAACCTTCTATGTGGCGCTCAACGCCAATGACCTTA 1960  
Oy 1862 AGCAGCCGGGTGCTTTGATGAGGTGCTCATCCGACATCAGTGAAGTACCTGGGACTGA 1921  
Db 1961 AGAGCTCGACAGTTTTCGATGAGGAGCGTGTGAGCAGCAGGTGCGATACCTTGGCCTAC 2020  
Oy 1922 TGGAGATCCTGCGGCGGAGAGCTGGCTTGGCTATGTTGGCAATTTGAGGCTTTC 1981  
Db 2021 TGGAGATCCTGCGGCGGAGAGCTGGCTTGGCTATGTTGGCAATTTGAGGCTTTC 2080  
Oy 1982 TGCAGAGTACAGTACATGCTGCTCCAGAGACATGGGCCATGTGGGAGAGACGCCACAG 2041  
Db 2081 TCGTGGCGTACAAAGATGATGCTGCACTACCTTGGCCCAATTTCCGTGGGGAGAGGACC 2140  
Oy 2042 ATGTGTGTGCGCTGTGTTGTCAGACACCTGGCTACAGCCAGAGAGTCAAAATGGGCA 2101  
Db 2141 GCGATGTGTGCTGCTGCTGATGAGGAAAGAAATTTCCCCAGGACGTTAAGTACGGCC 2200  
Oy 2102 GGACTAAGTCTTCATCCGATTCGCCAAGACCTTATTTGCCACAGAGGA 2150  
Db 2201 ACACAAAGATCTTTATACGCTCGCACGCACTGTGTTGCCCTGGAGCA 2249

RESULT 11  
AAS40995  
ID AAS40995 standard: cDNA: 3400 BP.  
XX  
AC AAS40995;

XX 17-DEC-2001 (first entry)  
DT  
XX  
DE cDNA encoding novel human enzyme polypeptide #211.  
XX  
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;  
KW ligase; hyperproliferative disorder; immunodeficiency disorder;  
KW autoimmune disorder; neurological disorder; metabolic disorder;  
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;  
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;  
KW anti arthritic; nephrotropic; anticoagulant; ss.  
OS Homo sapiens.  
XX  
PN WO20015301-A2.  
XX  
PD 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01239.  
PF  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225477.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.

```

PR 14-SEP-2000; 2000US-0233398.
PR 14-SEP-2000; 2000US-0233399.
PR 14-SEP-2000; 2000US-0233400.
PR 14-SEP-2000; 2000US-0233401.
PR 14-SEP-2000; 2000US-0233406.
PR 14-SEP-2000; 2000US-0233406.
PR 14-SEP-2000; 2000US-0233406.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234223.
PR 25-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234997.
PR 26-SEP-2000; 2000US-0234998.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0236337.
PR 29-SEP-2000; 2000US-0236337.
PR 29-SEP-2000; 2000US-0236337.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236368.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250161.
PR 05-DEC-2000; 2000US-0251030.

```

---

```

PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM,
XX WPI: 2001-465566/50.
XX P-PSDB; AA023125.
XX
XX Novel polypeptides and polynucleotides useful for diagnosing,
XX preventing, treating neural, immune system, muscular, reproductive,
XX pulmonary, cardiovascular, renal, proliferative disorders and cancers
XX diseases
XX
XX PS Claim 4; SEQ ID NO 221; 1180bp; English.
XX
XX The present invention relates to the isolation of novel human enzyme
XX polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences
XX encoding them. The enzyme polypeptides of the invention may comprise the
XX functional classes of oxidoreductases, transferases, hydrolases, lyases,
XX isomerases or ligases. The sequences of the invention are useful in the
XX diagnosis, treatment, prevention and/or prognosis of a wide range of
XX disorders, including hyperproliferative disorders (e.g. AIDS), cancer,
XX immunodeficiency disorders (e.g. AIDS), autoimmune disorders (e.g.
XX metabolic disorders (e.g. phenylketonuria), Alzheimer's disease),
XX (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
XX (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
XX blood-related disorders (e.g. haemophilia), reproductive disorders
XX (e.g. infertility) and infectious disorders (e.g. influenza). The
XX AAU0785-AAU1684 represent cDNA sequences encoding for the novel human
XX enzyme polypeptides of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pat_sequences.
XX
XX Sequence 3400 BP; 728 A; 1045 C; 1035 G; 591 T; 1 other:

```

---

```

Query Match 13.9%; Score 507.4; DB 22; Length 3400;
Best Local Similarity 54.7%; Pred. No. 3,3e-137;
Matches 1169; Conservative 1; Mismatches 917; Indels 52; Gaps 6;

```

QY	116	GACCTTGCTGCTGGAGAAATTTACACCACTGAGCGCTTCCTCATTTAGAACCTCCGGGG	175
DB	252	GACTTTTGCTTTTGGACCAAGTACCACTGAGGAG--GACTTCATGAGAACCTCCACCTC	308
QY	176	CGSTTCGGGAGAACCTCATTTATACCTACATGAGTCTGCTGCTCATGCTCTGTAATGCC	235
DB	309	AGSTTCGAGAAAGGCGCATCTACACCTACATGCTGAGTGCTGCTCGTGAACCC	368
QY	236	TACCGAGACCTACATGCTACAGCCGAGCATATGAAACCTACCTCGTGTGCTGCTGCT	295
DB	369	TACCGAGAGCTCCCTCTATGGGCTTATGGCCATTCGCGAGTACAGAGCCCTGACCTC	428
QY	296	TATGAATACACCTCATTTGTTGGATGGCTGACACTGATACCGGCACTTCTACT	355
DB	429	TATGAGCGGCAACCCCATCTATCTATGCTGCTGACCCCTCATCAAGCAATGAAGCAC	488
QY	356	GAGCTCGGAGACAGCAGTATTTCTGAGAGAGTGGGAGAGAGACAGAGAGCC	415
DB	489	CGGTCAGGAGACACCTGATCTGATCTCAAGGGAGAGTGGGAGAGAGAGAGAGAGCC	548
QY	416	ACCAAGAGACTGCTTCAGTTCTATGACAGAGACCTGCCAGCCCTGTAACGGGGTGGCGCA	475



Db 549 AGTAMGCATATGATGATATCGTCTGTCTGTCACCAATCCAAAGCCAGAGGCTGAGTGTG 608  
 QY 476 GTGCG-----AGACCGCTGTGGTAGAGCAACCCCGTGTAGAGGCTTTGGGAATGCC 529  
 Db 609 GAGAGGGTCAAGAGAGCTGTGCTCAAGTCCACCTGTGTCTGTGAGGCTTTGGCAATGGC 668  
 QY 530 AAGACTCTCCGACAGATATCTCCAGCCGGTTTGGAAAGTATCATGTGATGCTTTGAC 589  
 Db 669 CCGACCAACCGCAATCACAACCTCCACCCCTTTGGCAAGTATGACATCAACTTTGAC 728  
 QY 590 TTCAAGGGTGGCCCGTGGAGGCGCATTTCTGATTTACTCTGGAAAAGTCCCGGGTG 649  
 Db 729 TTCAAGGGGACCCGATCGAGAGACATCCACGCTACTGTGGAAGTCTCGGGTC 788  
 QY 650 GTGCACCAAAATTCAGGAGAGCGGAATTCACGCTTTTACAGCTATGAGGGGGC 709  
 Db 789 CTCAGCAGCAGCTGTGGTGAAGAATCTTCAGCCCTTTACCAATTTGCTGAGAGGCAT 848  
 QY 710 GAGGAGGAGACTCTCCGTGGCTGGCTTGGAAAGGACCCCGAGCTACTTTGACTTG 769  
 Db 849 GAGGACAAACACCTGCTCACTGACCTTGGAGAAACCTGCTGTATCAATTTTACA 908  
 QY 770 GTGAAGGGCCAGTGTGCCAAGGCT-----CCTCATTCAGACAGACAGATGAC 817  
 Db 909 CACCAAGGAGAGGACTCAACATGACTGTGCACAGTGCCTTGGACAGTATGAGAGAC 968  
 QY 818 TGAAGGTTATGAGAGGCGCTGTCCGTCATTTGACTCATGAGGATGAAGTGAAGAC 877  
 Db 969 CACAGGCGATGACGAGGCGCATGAGGCTCATGGCTTCAGTCTCAAGAGTGGAGTCT 1028  
 QY 878 TTGCTCAGCATGTGGCCAGCGTCTTACATCTGGSCAACATCCATTTGCTGTGACGAG 937  
 Db 1029 GTGCATCGCATCTGCTGCCATATTTGACCTGGAAACATCGAGTTTGTGGAGAGGAG 1088  
 QY 938 GACAGCAATGCCAGGTTACTACTGAGAACAGCTCAATATCTGACAGGCTCTTGGT 997  
 Db 1089 GAGGCTGGGCTGCTGAGAGAGGCGCTGGCAATGGCCGAGAGGCACTGTGTGACATGTG 1148  
 QY 998 GTGGAAGTACAAACA-----CTTAGGAACCCCTGACCCACAGAGAAG 1039  
 Db 1149 GCTGACTGACCGCCACACCCCGGAGCTGTGTCTGCTCCCTGCTGGCTGGCAGATT 1208  
 QY 1040 ATCATCGCCAAAGGGGAGAGCTCTGAGCCCACTGACCTTTGAAACAGCGGCTATGCA 1099  
 Db 1209 GCCTCGGAGGAGGAGAACTATAGAGAGGCGACACTGACAGTGGCCAGCTATGCC 1268  
 QY 1100 AAGGATGCCCTTGGCCAAAGGCTGTGTACAGCCGACATTTCACTGGCTGGTCAAGAAATC 1159  
 Db 1269 CGGGATGCTGTGCCAAGGAGTGTACAGCGGCTGTGTGAGTGGGTGGAACAGAGATC 1328  
 QY 1160 AATAGGTCACTGGCCCTTAAGGACGCTGAGAGCCGAGCTGGCGAAGCACACAGTTCTT 1219  
 Db 1329 AACAGTGTATGGAACCCCGGGGCCGGGATCCCTGCGGCTGATGGCAAGACAGTCAATT 1388  
 QY 1220 GGGCTCTTGACATTTAGCGCTTTGAAGTGTTCAGCATTAACAGCTTCAGACGTTCTGC 1279  
 Db 1389 GCGCTCTGTGACATCTATGAGCTTCGAGTGTTCCTCGTCAACAGTTTCGAGCGTTCTGC 1448  
 QY 1280 ATCAACTACTGCAATGAGAAAGCTGCAAGCTCTTATGAGTGTGACTGCTCAAGTGGAG 1339  
 Db 1449 ATCAACTACTGCAAAAGAGAGCTGACAGCTATTCATCAGCTCATCTCTGAGCGAA 1508  
 QY 1340 CAGGAGGAATAGAGGCTAGGCGATCGGCTGGGAACCTGTCCAGTACTTCAACACAAG 1399  
 Db 1509 CAGGAGGAATAGAGGCGGAGGCGATCCTGCAAGAGGCTTGAAGTATTTCAACAACGCC 1568  
 QY 1400 ATCATCTGTACCTGTAGAGAGAAATTCAAGGCAATCATCTTCTTGAAGAAG 1459  
 Db 1569 ACCATTTGTGATCTGTGAGGCGGCCACCGTGATCTCTGGCGTGGAGCGAGGCC 1628  
 QY 1460 TGGCTCGCTCTGGGAGGCGACAGGACCTGACTTTCTGGAAGAAGTTGGAGGACACTGTC 1519  
 Db 1629 TGGAGCTGTGCTGGACATCACTGACCGAATCTCTGAGACCCCTGGACGACGACAC 1688

QY 1520 AAGCCCCACCTTCACCTTCTGTACGACAAAGCTGCTGACCAGAAAGCAAGAAATCCCTA 1579  
 Db 1689 CGCATACCTACTACATACACACCGCCAGCTCTGCCCCACACAAAGACCATGAGTTT 1748  
 QY 1580 GACCGAGGGAGTTCGCCCTTCTGTCAATTATGCTGAGAGAGTGCATCACTGACTGGG 1639  
 Db 1749 GCGCGA---GACTTCCGGATCAAGACATATGACGGGAGCTCACATCTCCGTGGAAGGC 1805  
 QY 1640 TTTCGTGATTAACATATGACCTCTTCCGAACTGTAAGGAGAACATGTACAGCTCA 1599  
 Db 1806 TTATGACAAAGACAGAGATTCTTCTTCAGAGCTTCAAGCGCTGTCAACAGC 1865  
 QY 1700 ATGAACCCCATCAT-----GGCCAGTCTTTGACAAAGAGTGAAGCTCAGTCAAA 1749  
 Db 1866 ACGGACCCCACTCTAGGGGCGATGTGGCGGAGGCGAGAGACATCAACAGAGTGACC 1925  
 QY 1750 GAAGCGCCAGGACGCTGGCCACCCAGTTCAGATGAGCTCTGCAAGCTGTGAGATC 1809  
 Db 1926 AAGCGCCCTGACGGCTGGCACACTTTCMAAGACTCCATGTGGCCCTGTGGAGAAC 1985  
 QY 1810 CTGAGGCTTAGAGAGGCTGCTATATCCGGTGTATCAAGCCAAACGCGCAAGACGCC 1869  
 Db 1986 CTTCGCTCAAGAGGCTTCTTACGTCCGCTGCATCAAGCTCAATGAGCAAGTAGCT 2045  
 QY 1870 GGTGCTTTGATGAGGTGCTCATCCGACATCAGGTGAAGTACCTGGGACTGATGAGAAAT 1929  
 Db 2046 GGGAACTGGATGAGAAACACTGCTCGCACACAGTGGCATACCTGGGCGCTGGAGAAAT 2105  
 QY 1930 CTGCGCTGGCGAGAGCTGGCTTGTCTATGTGCGAAATATAGAGCTTTCTGACAGAG 1989  
 Db 2106 GTGAGGCTGCGAGGCGCTGCTTCCGCTCCGACCCCTACTCTGATTCCTGTACAG 2165  
 QY 1990 TACAACTCACTGCCCCAGACATGCGCCATGTGGGCAAGAGAGGCCCGAGATGGTGTG 2049  
 Db 2166 TACAAATGACCTGTGAATACATGAGCCCAACCACTGCTGGCTGCGCAAGAGCACC 2225  
 QY 2050 GCCGTTTGTGACAGACCTCGGCTACAAAGCCAGAAAGTGAATAATGGCAGAGACTAAG 2109  
 Db 2226 GTGAGGCTCTCTGAGAGAGCAGCGGCTGTCAGGGGAGCTGGCTTTGGCCACAGCAAG 2285  
 QY 2110 ATCTTATCCGATTTCCCAAGACCTTATTTGCCAAGAGAGACTCCCTGGAATCCGGCGG 2169  
 Db 2286 CTGTTCATCCGCTCACCCCGGACACTGTGCACACTGTGAGCAGAGCGAGCCGCTCATC 2345  
 QY 2170 CAGAGTCTAGCCACCAAGATCCAGCGGCGCTGGAGGGCC 2208  
 Db 2346 CCCATATTGTGCTGCTATTGTCAGAAAGGCAATGGCGGGC 2384

RESULT 12  
 AAS40998  
 ID AAS40998 standard; cDNA: 3332 BP.  
 XX  
 AC AAS40998:  
 AC  
 DT 17-DEC-2001 (first entry)  
 XX  
 DE cDNA encoding novel human enzyme polypeptide #214.  
 XX  
 KW Human: oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;  
 KW ligase; hyperproliferative disorder; immunodeficiency disorder;  
 KW autoimmune disorder; neurological disorder; metabolic disorder;  
 KW inflammatory disorder; cardiovascular disorder; reproductive disorder;  
 KW blood-related disorder; infectious disorder; gene therapy; cytostatic;  
 KW anti arthritic; nephrotropic; anticoagulant; ss.  
 XX  
 OS Homo sapiens.  
 OS  
 PN WO200155301-A2.  
 PN  
 PD 02-AUG-2001.  
 PD  
 XX



PT Novel polypeptides and polynucleotides useful for diagnosing,  
PT preventing, treating neural, immune system, muscular, reproductive,  
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous  
PT diseases

PS Claim 4: SEQ ID No 224; 1180bp; English.

XX The present invention relates to the isolation of novel human enzyme  
CC polypeptides (AA022915-AA023814), and the cDNA and genomic sequences  
CC encoding them. The enzyme polypeptides of the invention may comprise the  
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,  
CC isomerases or ligases. The sequences of the invention are useful in the  
CC diagnosis, treatment, prevention and/or prognosis of a wide range of  
CC disorders including hyperproliferative disorders (e.g. cancer),  
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders  
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),  
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders  
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),  
CC blood-related disorders (e.g. haemophilia), reproductive disorders  
CC (e.g. infertility) and infectious disorders (e.g. influenza). The  
CC polynucleotides of the invention can also be used in gene therapy.  
CC AA040785-AA041684 represent cDNA sequences encoding for the novel human  
CC enzyme polypeptides of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 3332 BP: 712 A; 1029 C; 1014 G; 577 T; 0 other;

Query Match 13.8%; Score 506.2; DB 22; Length 3332;

Best Local Similarity 54.7%; Pred. No. 7.2e-132;

Matches 1169; Conservative 0; Mismatches 918; Indels 52; Gaps 6;

OY 116 GACTTGTCTGCTGGAGATTTCACAGTGAAGGCTGCTTCAATGAGAACCTCCGGCG 175  
DB 195 GACTTTGTCTTTTGGACCAAGTACCATGAGAC---GACTTCAATGAGAACCTCCGAGCTC 251  
OY 176 CGGTTCCGGGAGAACCTTATATACCTACATCGTCTCTCTACTCTCTGTCAATCCC 235  
DB 252 AGGTTGAGAAAGGCGCGCATCTACACCTACATCGTGTGAGTGTGTCGCGTGAACCCC 311  
OY 236 TACCGAGACTTACAGATCTACAGCGGAGCATATGAAACGCTACCGTGTGTCAATTTC 295  
DB 312 TACCAAGACTCTGCTGATGAGGCTGAGGCGCATCGCAGGTGAGGCGCGTGAAGCTC 371  
OY 296 TATGAAGTACCACTTATTTTGCAGTGGCTGACATCTGTATACGGGCACTTCTGTACT 355  
DB 372 TATGAGCGGCGCAACCCATCTTATGCTGTGGCCACGCGCTTACAGGCAATGAAGCAC 431  
OY 356 GAGCGTGGGAGCCAGGACATGATGATTTCTGAGAGATGGGGCAGGACAGAGAGGCC 415  
DB 432 CGGTCCAGGAGACCTGATCGATCTCAGGAGGAGATGGGGCAGGAGAAACAGAGGCC 491  
OY 416 ACCAAGAGACTGCTCCAGTTCTATGAGAGACTGCCAGCCCTGAGACGGGTGGCGCA 475  
DB 492 ACTTAACCATATGATGATGATCTGCTGTATCCCAATTCAGGACGAGAGGTGAGGTG 551  
OY 476 GTGGC-----AGACCGCTGTGGAGAGACCCCGTGTGAGAGGCTTTGGAGATGGC 529  
DB 552 GAGAGGGTAAAGAGCTGTGCTCAAGTCCACCTGTGTCTGGAGGCTTTGGCAATGGC 611  
OY 530 AAGACTCTCCGCAACGATTAATCTACACCGGTTTGAAGATACATGATGTGCAATTGAC 589  
DB 612 CGCAGCAACCGCAATCACAACCTCAGCCGCTTTGGCAATGATGACATCAACTTTGAC 671  
OY 590 TTTCAGGGTGGCCCGGTGGAGGCGCAATCTTCAGTTACCTCTCTGGAAGAGTCCGGGTG 649  
DB 672 TTTCAGGGGAGCCCGATTCGAGAGACATTCACAGCTACTCTGGAAGAGTCTCGGCTC 731  
OY 650 GTGCACCAAAATTCAGGAGAGGAGGACCTTCAGCTTTTACAGTACTGTGAGGGGGG 709  
DB 732 CTCACAGCACGTTGGGTGAAGAACTTCACGCGCTTCTACCAATTTGCTGAGAGGCGAGT 791

OY 710 GAGAGAGACACTCTCCGTGGCTGGGCTTGGAAAGGACCCCGAGACTTGTACTCTG 769  
DB 792 GAGGACAAACAGCTGCTATGACTGACCTGTGAGAGAACTGCTGTATACATTTTCACA 851  
OY 770 GTGAAGGCGCAGTGTGCTCAAGGCT-----CCCTCATTCAGCAAGCAAGTGAAC 817  
DB 852 CACCAAGGAGACAGGACTCAACATGACTGTGCACAGTGCCTTGGACATGTGACAGAGC 911  
OY 818 TGAAGATTATGAGAGAGCGCTGCTCATTTGATGATCTACTGAGATGATGAGAGAGC 877  
DB 912 CACCAAGGAGATGACGAGGCGCATGAGGATCATGCGTTTCAAGAGTGTGAGTCT 971  
OY 878 TTGCTCAGCATGCTGGCCAGCGCTCTACATCTGCGCAACATCCACTTTCGCTGACAG 937  
DB 972 GTGCATCCGATCCGCTGCTCATATTTGACCTTGGAAACATTCAGATTTTGGAGACGAG 1031  
OY 938 GACAGCAATGCCAGGTTTACTGTGAAACACGCTCAATATGTGACAGAGCTCTTGGT 997  
DB 1032 GAGGTTGGGCTGCAGAAAGAGGCGCTGGCAGTGGCCGAGAGAGCACTGTGTGACATGTG 1091  
OY 998 GTGGAAGGTATCAACA-----CTTAGGGAAGCCCTGACCCACAGGAAG 1039  
DB 1092 GCTGACTGATGCGCCACACCCCGGAGCTGTGCTCCGCTCCCTGCTGTGCTGACAGTT 1151  
OY 1040 ATCATGCCAAGGGGGAAGAGCTCTGAGCCCACTGAACCTTGAACAGGCGCATATGCA 1099  
DB 1152 GCGTCGGAGGAGGAGGAGCACTATGAGAAAGGCGCACACTGTGACATGACGACATFGCC 1211  
OY 1100 AGGATGCGCTTGCCCAAGGCTGTACAGCCGACATTCACCTGTGCTGACAAAGATC 1159  
DB 1212 CGGGAAGCTGTGTCCAAGGAGGTATCCAGCGGCTTTGAGTGGGTGTGGAACAGAGATC 1271  
OY 1160 AATAGTCACTGGCCTCTTAAGAGAGCGCTGAGAGCCCGAGCGGCGAGACCAAGGTTCT 1219  
DB 1272 AACAGTGTATGAGAACCCCGGGGCGCGATCTCGCGCTGATGAGCAAGACAGATCAT 1331  
OY 1220 GGGCTCTGAGACATTTACGGCTTTGAAAGTTTTCACATTAACAGCTTCGAGAGTTCTGC 1279  
DB 1332 GGGCTGTGAGACATTTATGCTTTCGAGAGGTTTCCGCTGAAGGATTTTCGAGAGCTTGC 1391  
OY 1280 ATCACTACTGCAATGAGAAAGCTGCAGAGCTTCTATGAGAGTGTACTCTCAAGTGGAG 1339  
DB 1392 ATCACTACTGCAACGAGAGCTGCAGAGCTTATTCATCCAGCTCATCCGGAAGCGAA 1451  
OY 1340 CAGGAGAAATGAGGCTGATGGGATTCGCTGAGAACTGTCCAGACTTCAACAACAG 1399  
DB 1452 CAGGAGAGTACGAGCGCGAGGAGATCAGCTGACAGAGCTTATGATTTCAACAACGCC 1511  
OY 1400 ATCATCTGTACCTGTAGAGAGAGTTTCAGGGCATATCTCATCTTGTGATGAAGAG 1459  
DB 1512 ACCATGTGTGATGTGGTGGAGCGGCCCAACCGTGGCATCTGGCGTGGAGCGAGGCC 1571  
OY 1460 TGCCTGCTCTTGGGAGGCGACGAGCTGACCTTTCTGGAAGAGTTTGAAGAGACTGTG 1519  
DB 1572 TGCAGCTGTCTGGACCATCACTGACGCAATCTTCTGTGAGAGCCCTGTGAGACGACAC 1631  
OY 1520 AAGCCCAACCCCTGCTTCGAGCAGCAAGCTGCGTACACAGAAAGCCAGGAATCCCTA 1579  
DB 1632 CGCATACCTTACACTACACACGCGCCAGCTTGTCCCAAGACAAAGACATGAGAGTTT 1691  
OY 1580 GACCGAGGGGAGTTCCGCTTGTGATTTATGCTGAGAGGTGACATGATGTGACTGGG 1639  
DB 1692 GCGCGA---GACTTTCGAGATCAAGCACTATGAGGGAGAGTCAAGTACCTCGTGAAGGC 1748  
OY 1640 TTTCATGATTAAGCAATGACTCTCTTTCGGAACCTGAAAGAGACATGTGACAGTCA 1699  
DB 1749 TTTCATGAGCAAGAGAGAGATTTCTTTCAGAGCTTCAAGCGGCTGTGCTCAACAGC 1808  
OY 1700 ATGACCCCATCAT-----GGCCAGTGTCTTGAACAGAGTGTGATGACAA 1749  
DB 1809 ACGGACCCCACTTACGGGCGATGTGGCGGAGCGGACGAGCAATCAAGAGAGTGAACC 1868  
OY 1750 GAAGCGGCGAGAGGAGTGGCCACCCAGTTCAAGATGAGCTCTGTGAGAGTCTGTGAGATC 1809

```

Db 1869 AAGCCGCCCCCTGACGGCTGGCAGACACTCTTCAAGAACTCCATGTGGCCCTGTGGAGAAC 1928
Oy 1810 CTGAGGCTTAAGAGAGCTCCCTATATCCGGTGCATCAAGCCAAAGCCCAAGCAGCGC 1869
Db 1929 CTTCGCTCCAGAGAGAGCCCTTCTAGTCCCTGCTCATCAAGCCCAATGAGACAAAGTAGCT 1988
Oy 1870 GGTGCTCTTTGATGAGTGTCTCATCCGACATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 1929
Db 1989 GGGAGAGCTGATGAGAAACACTGTCCGCAACGACCTGATACCTGGGCTGCTGGAGAAAT 2048
Oy 1930 CTGGCGCTGGCAGAGACTGGCTTGGCTTTCGTCGAATATGAGCTTTCCTGAGAGG 1989
Db 2049 GTAGGGCTCGCAGAGCTGGCTTGGCTTTCGTCGAATATGAGCTTTCCTGAGAGG 2108
Oy 1990 TAAAGTCACTGTGCCCCAGACATGGCCATGTGGGAGAGAGCCCAAGATGTGTG 2049
Db 2109 TACAAATGACCTGTGAATATACATGAGCCCAACCACTGTGGGCTCCGACAAAGCAGCC 2168
Oy 2050 GCGGTGTGTGTCAGACACCTGGCTACAAAGCAAGATGACAAATGAGCAGAGACTAAG 2109
Db 2169 GTAGAGCTCTCTCTGAGCAGACAGCGGCTGAGAGGAGAGCTTGGCTTGGCCACAGCAAG 2228
Oy 2110 ATCTTATCCGATTCCTCCCAAGACCTTATTTGACACAGAGAGAGCTCCCTGGAATCCGCGG 2169
Db 2229 CTGTTATCCCTCTACACCCGAGACACTGTGTCACACTGAGCAGACAGCCGCTCTCATC 2288
Oy 2170 CAGACTTACGACCAAGATGACAGCGGCTGAGAGGCGC 2208
Db 2289 CCATCATTTGCTGCTGCTATTTGCAAGAGCATGGGCGGCG 2327

```

## RESULT 13

ABK52205  
ID ABK52205 standard; DNA; 3617 BP.

AC ABK52205:

DT 13-AUG-2002 (first entry)

DE DNA encoding human Myosin-1F protein MYO1F.

KM Human: phosphoinositide-binding protein; phosphatidyl acid; gene; ds;  
KM phosphatidic acid; signal transduction; housekeeping; myosin-1F; MYO1F.  
OS Homo sapiens.

EH Key Location/Qualifiers  
FT CDS 41..3337  
FT /\*tag= a  
FT /product= "Human myosin-1F protein MYO1F"

PN W0200218946-A2.

PD 07-MAR-2002.

PP 23-AUG-2001; 2001MO-GB03791.

PR 23-AUG-2000; 2000GB-0020833.

PR 15-DEC-2000; 2000GB-0030637.

PA (BABR-) BABRAHAM INST.  
PA (UYCA-) UNIV CAMBRIDGE TECH SERVICES LTD.  
XX (LIMZ/) LIM Z.

PI Stephens L, Hawkins PT, Holmes AB, Manfava M, Ktistakis N;  
PI Thuring JWF;

DR WPT: 2002-434908/46.  
DR P-PSDB: AAU97544.

PT Novel probe for use in assay method for detecting, measuring,  
Identifying and/or isolating PA- and/or PIPn-binding protein in a test

```

PT sample, has immobilised phosphatidyl acid attached to solid support
XX
XX Disclosure: Page 133-134; 163pp; English.
PS
XX The present invention relates to a new probe comprising or consisting of
XX an immobilised phosphatidyl acid derivative attached on to a solid
XX support, or a phosphatidic acid (PA) functionalised solid support. The
XX probe of the invention is useful in an assay method for identifying
XX and/or isolating a protein that binds to the probe. The invention is also
XX useful for detecting, measuring, identifying and/or isolating more than
XX one type of phosphatidic acid and/or phosphoinositide-binding protein
XX (PIPn) from a test sample e.g. a tissue of tissue culture extract. The
XX assay is also useful for detecting, measuring, identifying and/or
XX isolating phosphatidic acid and/or phosphoinositide-binding and/or
XX a test sample, to detect and/or measure the ability of an agent, applied
XX to phosphatidic acid and/or phosphoinositide-binding protein-containing
XX test sample, to agonise or antagonise protein-probe binding, and to
XX agonise or antagonise the ability of an agent, applied to the probe, to
XX identify an agonist or antagonist of phosphatidic acid/
XX phosphoinositide-binding protein-phosphatidic acid/ phosphoinositide
XX interaction, and in a single step high throughput screen of candidate
XX agonist and/or antagonist. The invention is also useful for identifying
XX important proteins for signal transduction, housekeeping and diagnosis.
XX The probe is useful as an important research tool in fundamental research
XX for diagnostics and drug discovery. The present nucleic acid sequence
XX encodes the human myosin-1F protein MYO1F of the invention.
SQ
Sequence 3617 BP; 809 A; 1103 C; 1073 G; 631 T; 1 other;

```

## Query Match

Best Local Similarity 12.6%; Score 460.8; DB 24; Length 3617;  
Matches 1149; Conservative 0; Mismatches 857; Indels 94; Gaps 8;

```

Oy 140 ACCAGTGAAGGCTGCTTATGAGACCTCCGCGGGGCTTCCGAGAGAACCTCATTTAT 199
Db 122 ATCACCAGGAAGCCGATTTGCGCCACCTCCGGAAGGCTTATGAGACCTCATTTAT 181
Oy 200 ACCTCATTCGCTCTGCTAGTGTCTGTCATTCCTCAGAGACCTCATGATTTACAGC 259
Db 182 ACCTCATTCGCTCTGCTAGTGTCTGTCATTCCTCAGAGACCTCATGATTTACAGC 241
Oy 260 CGGAGCATATGAAAGCTTACCGGCTGAGGCTGATTTATGAAGTACCACTCATTTGTT 319
Db 242 GACCGTGAATCAGACCTTATCAGGCGCGGCCCAATGATGAAGATCTCCGACATCTAC 301
Oy 320 GCACTGCTGACACTGATATACCGGCACTTCTGACTGAGAGCGGAGACAGCAGTATG 379
Db 302 GCCCTCAGGAGCAACATGTAACCGAATCTTATCAGCTGTGAGAACCACTGTGTATC 361
Oy 380 ATTTCTGAGAGAGTGGGAGGAGCAAGAGAGAGGCGACCAAGATATATCATGGGTATC 421
Db 362 ATTAGTGAAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 439
Oy 440 GCAGAGACCTGCCCAAGCCCTGAAACGGGGTGGCGAGTGGAGAGAGAGAGAGAGAG 499
Db 422 TCCAAAGTGTCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 481
Oy 500 AACCCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 559
Db 482 AACCCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 541
Oy 560 TTTGGAAGTACATGATGTGACATTTGACTTCAAGAGTGGCCCGCTGGAGAGAGAGAG 619
Db 542 TTTGGAAGTACATTTGAGATTCAGCTCAGCGAGTGGAGAGAGAGAGAGAGAGAGAG 601
Oy 620 CTCAGTTACCTCTGGAAGAGTCCCGGAGTGTGCGACCAATACAGAGAGAGAGAGAG 679
Db 602 TCCAAATCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 661
Oy 680 CAGCTCTTTACAGACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 719
Db 662 CACATCTACTACAGCTGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 721

```

OY	740	GAACGGAACCCCCAGAGACTCTGTGACCTGCGTGGGAAGGGGCAATGTGCCAAAGCTCTCCMC	799
Db	722	AT---GACACCGGACTACTACTTACTTACTACTCTCAACCAATGGAACACTACCAAGTGGAGGCG	778
OY	800	ATCAACACAAGAGTGAAGGTTATGAGGAAGGGCGCTGTCCGTATTGACTTCACT	859
Db	779	ACGGACGACAGAAAGCGCTTGGTGTGACTGTGAGCGTATGACAGTTATTGGATCCCG	838
OY	860	GAGGATGAATGGAAGACTGCTCAGCATGCTGGCCAGCGTCTTACATCTGGGCAACATC	919
Db	839	CCGAGACTCCAGGAGGTGGCTGTCCAGCTCGTGGGGGATCTTGGACCTGGGGAACATC	898
OY	920	CACTTTCTCTGCTGACAGGACAGCACTGATCCGAGTTACTAGAGAACCACTCAAAATAT	979
Db	899	AGTTTCTGTGAACGCG---AATTACGCCCAAGTGGAGAGTGTGGAGACTCTTGCGCTTT	955
OY	980	CTGACCAAGCTCTTGGTGTGGAAGGTACACACTTGGGAAGCCCTGACCCACAGGAG	1039
Db	956	CCCCCTTACCCTGTGGGCAATTGACAGCGGGGCACTGAGGAAMACATGACACGCGCAAG	1015
OY	1040	ATCATGCCCAAGGGGGGAAGGC-----TCGTAGCCCACTGACCTTGACAG	1087
Db	1016	ATGGACAGCCGCTGGGGCGGGCCCAAGCACTCATCATGTGACCTTAAGGTGAAGAG	1075
OY	1088	CGCGCATATGCAAGGGATGCGCTTGGCCAAGGCTGTGTACAGCCGACATTCACCTGGCTG	1147
Db	1076	GCACCTTACACCCGTGTATGCTCGGCCAGAGGGGCTGTATGCCCGCTCTGTGACTTCCTC	1135
OY	1148	GTCAGAAAGATCAATAGTGTACTGCGGCTCTTAAGAGCGCTGAGAGGCCCACTGGCGAAGC	1207
Db	1136	GTTGAGGCGCCATCAACCGTGTATGCAAGAAACCCACAGGAAGATTAAGATC-----	1186
OY	1208	ACCAAGCTTTTGGGGCTCCGAGCAATTATAGGCTTTGAAGTGTTCACATACAGCTTC	1267
Db	1187	-----GGTGTGCTGGACATTTACGGCTTGGANATCTTCGAAAMAAATGSGCTT	1234
OY	1268	GAGCAGTTCTGCATCAACTACCTGCAATGAAAGCTGACAGCTCTTCATGAGCTGACT	1327
Db	1235	GAGCAGTTTGTGCACTCAACTTCGTCAATGTGAAGAGCTGACCAATCTTATGTGAACCTTACC	1294
OY	1328	CTCAAGTCGGAAGGAGGAGAAATGCGAGGCTGAGGGCATTCGCTGGGAACCTGTCCAGTAC	1387
Db	1295	CTGAAAGCCGACGAGGGAGGTATGTGACAGGAAGGATTCGCTGGACATTCACATCCAGTAC	1354
OY	1388	TTCAACAACAAGATCACTCTGTGACTCGTGTAGAGAGGAAGTTCA-----GGGCAATCATC	1441
Db	1355	TTCAACAACAAGGTCGTGTGACTCATTCAGAAACAAAGCTGAGCCGCCCAAGCATATG	1414
OY	1442	TTCACTTTTGGATGAAGTGC-----CTGCGTCTGGGGAGGCCACAGCACTGTAC	1492
Db	1415	AGCGCTTTGACAGCGTGTCCGACCATGACAGCGCACGGGGGGGAGACAGCAGCA	1474
OY	1493	TTTTGTGAGAAGTTGGAGACATCTGTCAAGCCCAACCCCTCACTTCTGACGACACAAGTCT	1552
Db	1475	CTGCTGGAGAAGGTCGAGGGGCTGTGGGGGACCAAGACAACTTTCAAAGCTGGA-----	1529
OY	1553	GCTGACCAAGAAACAGGAATCCCTAGACCGAGGGGAGTTCGCGCTTCTGCATTATGCT	1612
Db	1530	-----GCGCGCGCTTGTGTATCCACACACTACCT	1558
OY	1613	GGAGAGGTGACTTACAGTGTGACTGGGTTTGTGSAATAAACAAATGACCTCTCTGTCCG	1672
Db	1559	GGCAAGGTTCTCTACAGCTCACGGGCTTCTGCGAAGGAACCGAGAGCTTCTCTTCC	1618
OY	1673	AACCTGAAGAGACATGTGCACTCAATGAACCCCATCATGGCCCAAGTGTCTGTGACAG	1732
Db	1619	GACCTCTAGAGGTGAATGACAGACCAAGTGAAGAGGCTTCTCGGATGCTGTCTCCGAG	1678
OY	1733	AGTAGCTCACTGACAAAGAGCG---CCAGAGCGTGTGGCACCACATTCAAGATGAC	1788
Db	1679	AAGCTGATGGAGCAAGAGGGGCGCCCAAGACCGCGCGGTCCAAATCAAGAAACAA	1738

QY	1789	CTTGTGAGCTCGTGGAGATTCCTGAGAGCTTTAAGGAGCTGCCTATATCCGGTGCATCAAG	1848
Db	1739	GCCACGACCTGTGTGGCCACTGATGAGGTGCACACCCCACTACATCCGTGCATCAAA	1798
QY	1849	CCAAATCAGCCCAAGCAGCCGGTGCCTTTGATGTAGTGTCCTATCCGACATCAGGTGAG	1908
Db	1799	CCCAAGCAGACCAAGCACCGCCGAGACTGGAGAGAGAAACAGATCAACACACAGGTGGA	1858
QY	1909	TACCTGGAGCTGATGAGAAATTCGCGGTGGCAGAGCTGCGCTTATCGTCGAAA	1968
Db	1859	TACCTGGGCTGAAGAGAAATCATCAGGGTGGCCAGAGCGCGCTTGCCCTACCGCGCCAG	1918
QY	1969	TATGAGGCTTCTCGCAGAGGTCAAGTACGTGTGCGCCAGAGACATGGCCCATGTGGGCA	2028
Db	1919	TTTCCCAAAATTCCTCGAGAGGTATGCCATTGTACCCCGAAGACCTGGCCGCGGGCGGT	1978
QY	2029	GGAAGCGCCCAAGAGATGGTGTGGCCGTGTGGTCAGACACTCGGTGATACAGGCAAGAAAG	2088
Db	1979	GGGAGCAGACGCCAGAGGCGTTCAGACACTGTCTTGGGGCGGTATACATGAGAGCCGACCG	2038
QY	2089	TACAAAATGGGCGAGACTTAAGATCTTTCATTCGATTCGATTTCCCAAGACCTTATTTGCCACAG	2148
Db	2039	TACCAATGGGGGAGACACCAAGGTCTTTGTCAAGAACCCAGATGCTCTTTCCTCTGGAG	2098
QY	2149	GACTCCCTGGAATCCGGCGGCGAGAGTCTAGCCACCAAGATCCAGCGCGCTGTGAGGGGC	2208
Db	2099	GAGGTGGGAAGGCAAAAGTTGATGTGCTTTGGCCCAACATCCCAAGAGCTGTGGCGGCGC	2158

RESULT 14  
AAFC32774/C  
ID AAF32774 standard; cDNA; 2263 BP

AC AAF32774;

DT 22-MAR-2001 (first entry)

DE Human secreted protein cDNA #18.

KM Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KM antiallergic; hepatotropic; antidiabetic; antiinflammatory; antileuer;  
KM vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KM cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder  
KM neurological disease; infection; human; secreted protein; ss. .

05 Homo sapiens.

PN WO2000077197-A1

PD 21-DEC-2000.

PE 01-JUN-2000; 2000WO-US14934.

PR 11-JUN-1999; 99US-0138599.

PA (HUMA-) HUMAN GENOME SCI INC.

XX

XX

DR P-PSDB; AAB64566.

PT Isolated nucleic acid molecule encoding a human secreted protein is

XX

XX

CC 47 huna secreted proteins ABB6549-B64594. The genes can be used to  
CC generate fusion proteins by linking to the gene for the human  
CC immunoglobulin G Fc portion (SEID1) for increasing the stability of  
CC the fusion protein as compared to the human protein only. The genes and

CC proteins are useful for preventing, ameliorating or treating medical  
 CC conditions, e.g. by protein or gene therapy. The genes are isolated  
 CC from a range of human tissues disclosed in the specification. The  
 CC nucleic acids, proteins, antibodies and (ant)agonists are useful in  
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischemia;  
 CC wound healing; (e) neurological diseases such as cerebral ischemia;  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.

SO Sequence 2263 BP; 563 A; 529 C; 459 G; 712 T; 0 other;

Query Match Best Local Similarity 12.3%; Score 450.6; DB 22; Length 2263;  
 Matches 1083; Conservative 0; Mismatches 909; Indels 37; Gaps 5;

142 CAGTACGCTGCTTCATTTGAGACCTCCGCGCGCTTCGCGAGACCTCATTTATAC 201  
 2016 CAAATAGAGACCTTTCATCAACACCTCAAGAGCGCTTGAACACAGTGAATATACAC 1957  
 202 CTACATCGCTCTGCTCTCTGTCATCTCCCTACCGAGACCTACAGATCTACAGCG 261  
 1956 TTACATTGGAAGTGTGTTATATCTGTAAACCATATGCGTCTTACCATTTATTCACC 1897  
 262 GCACATATGGAACCTACCGTGTGTGCTTCTATATAGTACCTCATTTGTTTGC 321  
 1896 AGAAGAAAGTGAAGAAATACAGACACAAATTTTATGACCTGAGCCCTCACATCTTGC 1837  
 322 AGTGGCTGACACTGTATACCGGGCACTGTACTGAGCGTGCAGGACGACGATGAT 381  
 1836 CCTTTCGAGTGAAGATACAGATCCCTTACGATCAAGATTAAGACCATATATCTCAT 1777  
 382 TTCTGAGAGAGTGGGCGACGACCAAGACAGAGCCACCAAGACTCTCTCACTTATAC 441  
 1776 TACTGGGAAGTGAAGTGAAGAAACAGAGGCGCATAGCTTGTATGCTTATGTCG 1717  
 442 AGAGACCTGCCAGCCCTGACGCGGTGCGGACGTCGACAGCCGCTGTGCGAGGAA 501  
 1716 AGCTGTGTTGGAAAGAGACAGATTAATCAAGTTAAAGTAAAGAGCTTTTACAGTCCA 1657  
 502 CCCCGTGTAGAGCGCTTGGGAAATCCCAAGACTCTCCGCAAGTAACTCCAGCCGTT 561  
 1656 CCCGCTCTGGAAGCTTTTGAATGCCAACAACCTGAAGAAATGACAACCTCTCTGAT 1597  
 562 TGGAAAGTACATGATGTGCACTTGTCAAGAGTGCCTCCGCTGGAGGACCATATTC 621  
 1596 TGGCAATATATGATATTTGAATTTGACTTAAAGCGATCCACTGAGGAGATATAG 1537  
 622 CAGTTACCTCTGGAAGATCCGCGGTGTGACCAAAATCAAGAGACGCGGACCTTCCA 681  
 1536 TACTATCTTTTGAAGAAATCTCGGTTTAAACAGCCCAAGAGTGAAGAAACTTTCA 1477  
 682 CCTCTTTTACAGCTGAGGAGGCGGAGAGAGACTCTCCGCTGGCTGGCTTGA 741  
 1476 TGTGTTTATGAGCTCTCTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1417  
 742 ACGAAGCCCAAGAGTACTTGTACTGTGTGAAGGCGCAGTGTGCCAAGTCTCTCAT 801  
 1416 GAGGAGATTTTACAGCATATACCTAGCTGTG---ATTGCGCCCAAGATGAGAT 1360  
 802 CAACGACAGAGTGTGAGAGTATATAGAGAGCGCTGTCCCTATTTACTTCACTGA 861  
 1359 GGTATGATGACGCAAAATTTTGAACCCGTGCGGAATGCCATGACAGATTGTGGCTTATGA 1300  
 862 GGATGAAGTGAAGACTGTCTACAGATCTGTGGCCAGGCTCTACATCTGTGGACATCA 921  
 1299 TCATGAAGCTGAGTCTGTCTTGGCGGTGTGTGTCAGCAGTGTGTGAACCTGGGACATTTGA 1240

922 CTTTGTGCTGACGAGACACACATATGCCAGTTACTACTGAGAACGACTCAA----- 976  
 1239 GTTCAACCCGAACTCTGAGTGAATGCTTATGTAAGAACCAATCAAAATTAAGAG 1180  
 977 -TATGACCAAGCTCTTGTGTGAGAGTACAAACACTTATAGGAGAGCCCTGACCCACAG 1035  
 1179 TTAAGAAATTTGTGAATTTGCCGCTATGATCAATCACTGTAGACGATTCAGTT 1120  
 1036 GAACATATCTGCCAAGAGGAGAGCTGTGAGCCCACTGAAACCTTGAAGCGGCGAT 1095  
 1119 CCGAACAGTTGAGCCAAACAGAAAGTTTCACTACACTGAAATGTGCTCAGGCTTATTA 1060  
 1096 TGCAGGAGATGCGCTTGGCCAAAGCTGTGTACGCGGAGACATTCACCGCTGTGAGAA 1155  
 1059 TGCCCGTATGCTGTGCTGCTTAAACACCTCTACAGAGGTTGTTTCAATGTTGTAAATCG 1000  
 1156 GATCAATATGATCATGCGCTCTTGAAGACGCGGAGACCCAGCTGAGCAGCCAGCT 1215  
 999 AATCAATG-----AAACATTAAGGCAACAAACAAAGTGAAGAAAGT 955  
 1216 TCTTGGCTCTCGACATTTAAGCTTTTGAAGTGTTCAGATTAACACTCTGAGCAT 1275  
 954 CATGGTGTCTGTGACATTTATGCTTTGAGATTTTTCAGAGCAACAGCTTTGAGCAGTT 895  
 1276 CTGCATCACTACTGCAATGAGAGAGTGCAGAGCTCTTCATTCAGCTGACTCTCAAGTC 1335  
 894 CATTATTAATTAATTTATGAGAAAGCTGCAACAAATCTTCACTGAACTCTTAAAGA 835  
 1336 GAGCAGAGAGAAATACAGAGGAGAGGAGGATGAGTGTGAGACCTGTCAAGTCAACAA 1395  
 834 AGGCGAGAGAGATATTAACGAGAGATATGAAATGAGTCACTGACTACTTCAATTA 775  
 1396 CAAGATCATCTGTGACCTGTGAGAGAGAAATCAAGGAGATCTCATCTCTGATGA 1455  
 774 TCTATCATTTGTGCTGACCTTAATTAAGAAATTAACAAATGATCTGCGCTGAGATGA 715  
 1456 AGAGTCTGTGCTGTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 715  
 714 AGAGTCTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 715  
 1516 TGTCAAGCCCACTCTACTT-----CTGACGACAGAGCTCTGACCAAGAAAGCAG 1569  
 654 ATGTGCCCAACACAGCATTTTGAAGAGATGAGAAAGTGTCTGTGCTCTCATAGA 595  
 1570 GAATCCCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1629  
 594 CAGCTCTGTGCTGACAGCTGTCTGATGATCCAGCATTTATGCTGTGAGAGGAGGAGGAG 535  
 1630 TGTGACTGTGCTGTGATTAAGAAATGAGCTCTCTCTGAGAACTGAAAGAGGAGGAG 1689  
 534 GTTGAAGGATGCTGTGACAAAGAAATGATGATGATGATGATGATGATGATGATGATGAT 475  
 1690 GTTGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1749  
 474 GTTGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 415  
 1750 GAAGGAGGAGGAG-----AGGTGAGCAGGAGTGAAGATGAGGAGGAGGAGGAGGAGGAG 1802  
 414 CAACCTTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 355  
 1803 GAGATCTGTGAGTGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1862  
 354 GAAAGAACTTACAGAGCAAGAAAGAACTTAATTAATTAATTAATTAATTAATTAATTAAT 295  
 1863 GCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1922  
 294 AGCAGACACATCTTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 235  
 1923 GAGAGATCTGTGAGTGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 235  
 234 GAGAGATCTGTGAGTGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 175  
 1983 GCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2042





Db 1577 TGAATATTTCCAGAAAAATGCTTTGAAACAGTTTTGTATCAATTTTGTATGAAGAACT 1636  
QY 1303 GCAGCAGCTCTTTCATCGACCTGACTCTCAAGTCGAGCAGAGAGAAATACGAGCCTGAGGG 1362  
Db 1637 GCAGCAGATTTTATTTGAACTTACATTAAGCAGAACAGAGAAATATGTTCAAGAGGG 1696  
QY 1363 CATCGCGTGGAGCTGTCCACTTCTACACAAAGATCATCTGTGACCTCGGTAGAGGA 1422  
Db 1697 AATTAAGATGACACCCATTGAGTACTTAAATAAATCGTTATGTGACCTCATAGGAA 1756  
QY 1423 GAAGTTCAA-----GGCATCATCTCCATCTTGGATGAAGAGTCCCTGCTCGGGA 1476  
Db 1757 CAAAGTGAACCTCTCTGAGCATATGAGCATCTGATGACGTCTCCAGAAACTTCAGAT 1802  
QY 1477 GGCACAGGACCTGACTTCTTGAGAGATTGAGGACATGTCAGAGCCACCTCACTT 1536  
Db 1803 -GCCACGATGCTCGGTGGGTGAGGGGAGATCAGACGCTGCTCCAGAACTTCAGAT 1861  
QY 1537 CCTGACGACAAAGCTGCTGCTGACCGAAGACAGAAATCCCTAGACCGAGGGAGTCCG 1596  
Db 1862 GCAGATTGGAGTCATGAGCACTTCAAGTGGAACTCA-----AGGCTTCAT 1909  
QY 1597 CTTTCTCATTTATGCTGAGAGCTGACTACAGTGTGACTGGGTTTCTGTGATAAACA 1656  
Db 1910 CATTCATCATTTATGCTGGAGAGGTATCTATGACATGATGGCTTTGTGAAGGAAACG 1969  
QY 1657 TGACCTCTCTCTCGGAACTGAGAGACATGTGACAGCTCAATGAACCCCATCATGGC 1716  
Db 1970 GGAATGCTTTTATGATCTCTATGAGCTTATGAGAGCAGACGAGCTGCTTTCATAA 2029  
QY 1717 CCAGTCTTTGACAAAGTGAAGTGAAGTGAAGAGGCG---CCAGAGCGGTGGCCAC 1772  
Db 2030 GTCTTTATTTCCGAAAAATCTGAGAGCTGACAGAAAGGCGCCCACTACTGCGGAG 2089  
QY 1773 CCAGTTCAAGATGAGCTCTCTGACAGTCTGAGATCCTGAGGTCTAAGAGCCTGCTA 1832  
Db 2090 CAAATTAAGAAACAAGCCATGACCTTGTGAGCACCTGATGAATGTACGCCACTA 2149  
QY 1833 TATCCGCTGATCAAGCCAAAGCAGCCAGCAGCGGCTGCTTTGATGAGGTGCTCAT 1892  
Db 2150 CATTGCTGATCAAGCCAAAGCAGAACCAAGACCTGGAGAGAAAGCAAGGT 2209  
QY 1893 CCGACATCAGGTGAAGTCTGAGAGTGAAGATCTGCGCTGCGCAGAGCTGCTT 1952  
Db 2210 AAGCATCAAGTCAATATTTTGGGTCTGAAGAGAACATTCGAGTGAAGAGACTGCTA 2269  
QY 1953 TGCCTATGCTGCAAAATATGAGCTTCTGAGAGGTACAGTCACTGTGCCAGAGAC 2012  
Db 2270 TGCCTATGCGGCAATCTTCCAAAAATTCCTACAGAGGTATGCCATTCAGCAAAAGCCAC 2329  
QY 2013 ATGCCCCATGTGGGAGAGGCGCCCAAGATGCTGCGCTGTGTCAGACACCTCGG 2072  
Db 2330 CTGGCCTTTTGGAGGAGAGAGAAAGCAAGGCTCTGACAGCTGCTGACATCGGTCAA 2389  
QY 2073 CTACAAAGCAGAGATACAAATGGGAGAGTATGATCTTCAATCCGATTTCCCAAGAC 2132  
Db 2390 CATGAGACGAGCAGATTCAGCTGGGAGAGTAAAGTGTTCATCAAGCCCGAGCTC 2449  
QY 2133 CTTATTTGCCACAGAGACTCCCTGGAATCCGGCGCAGAGTCTAGCCACCAGATGCA 2192  
Db 2450 TCTATTTCTTTTGAAGAGATGAGAGAAAGTATGATGGTATGCTCGAGTATATCA 2509  
QY 2193 GCGGCGCTGAGAGGCTTTTCAATTGGCAGAGAAATTTCCCGGCTGGAAGCGATCAGCC 2250  
Db 2510 GAATTCATGAGGAAATGTGTGGCCGGAAGAAATACGTTCAATGAGAGAAAGCC 2567

Search completed: December 9, 2002, 12:33:02  
Job time : 548 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: December 9, 2002, 12:22:35 ; Search time 91 Seconds

(without alignments)  
12337.842 Million cell updates/sec

Title: US-09-893-371-3

Perfect score: 3661

Sequence: 1 ggagcggggcgccggtccg.....ccacctgacctttagcca 3661

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Issued\_Patents\_NA:\*

- 1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	149.4	4.1	5661	US-08-938-105-2	Sequence 2, Appli
2	104.6	2.9	8473	US-09-172-422-2	Sequence 2, Appli
3	80.2	2.2	693	US-08-998-416-1135	Sequence 1135, Ap
4	43.2	1.2	364	US-08-905-223-162	Sequence 162, App
5	40.4	1.1	343	US-08-473-020A-6	Sequence 6, Appli
6	39.2	1.1	1208	US-08-403-852D-4	Sequence 4, Appli
7	39.2	1.1	1208	US-08-510-646B-4	Sequence 4, Appli
8	39.2	1.1	1208	US-09-231-818-4	Sequence 4, Appli
9	39.2	1.1	5392	US-08-403-852D-1	Sequence 1, Appli
10	39.2	1.1	5392	US-08-510-646B-1	Sequence 1, Appli
11	39.2	1.1	5392	US-09-231-818-1	Sequence 1, Appli
12	39.2	1.1	4411529	US-09-103-840A-1	Sequence 1, Appli
13	39	1.1	1664	US-09-339-993-1	Sequence 1, Appli
14	38.6	1.1	1399	US-08-471-033-24	Sequence 24, Appl
15	38.6	1.1	1399	US-08-471-044-24	Sequence 24, Appl
16	38.6	1.1	1399	US-08-463-483A-24	Sequence 24, Appl
17	38.6	1.1	1399	US-08-471-046A-24	Sequence 24, Appl
18	38.6	1.1	1399	US-08-470-566B-24	Sequence 24, Appl
19	38.6	1.1	1399	US-08-469-33A-24	Sequence 24, Appl
20	38.6	1.1	1399	US-09-300-529-24	Sequence 24, Appl
21	38.4	1.0	30001	US-08-125-468-1	Sequence 1, Appli
22	38.4	1.0	30001	US-08-474-933-1	Sequence 1, Appli
23	37.8	1.0	957	US-07-745-206A-16	Sequence 16, Appl
24	37.8	1.0	957	US-08-311-363-16	Sequence 16, Appl
25	37.8	1.0	5467	US-07-745-206A-12	Sequence 12, Appl
26	37.8	1.0	5467	US-08-311-363-12	Sequence 12, Appl
27	37.8	1.0	6232	US-08-456-200B-11	Sequence 11, Appl

28	37.8	1.0	7175	1	US-08-455-543A-8	Sequence 8, Appli
29	37.8	1.0	7175	2	US-08-193-078B-8	Sequence 8, Appli
30	37.8	1.0	7175	2	US-08-223-305C-8	Sequence 8, Appli
31	37.8	1.0	7175	2	US-08-149-097D-8	Sequence 8, Appli
32	37.8	1.0	7175	3	US-08-949-386-8	Sequence 8, Appli
33	37.8	1.0	7175	3	US-08-450-563-8	Sequence 8, Appli
34	37.8	1.0	7175	4	US-08-984-709A-8	Sequence 8, Appli
35	37.8	1.0	7175	4	US-08-450-272-8	Sequence 8, Appli
36	37.8	1.0	7177	4	US-09-268-163-7	Sequence 1, Appli
37	37.8	1.0	7266	3	US-08-713-118-1	Sequence 1, Appli
38	37.8	1.0	7266	3	US-09-452-007-1	Sequence 1, Appli
39	37.8	1.0	7362	1	US-08-455-543A-7	Sequence 7, Appli
40	37.8	1.0	7362	2	US-08-193-078B-7	Sequence 7, Appli
41	37.8	1.0	7362	2	US-08-223-305C-7	Sequence 7, Appli
42	37.8	1.0	7362	2	US-08-149-097D-7	Sequence 7, Appli
43	37.8	1.0	7362	3	US-08-949-386-7	Sequence 7, Appli
44	37.8	1.0	7362	3	US-08-450-562-7	Sequence 7, Appli
45	37.8	1.0	7362	4	US-08-984-709A-7	Sequence 7, Appli

## ALIGNMENTS

RESULT 1  
US-08-938-105-2  
; Sequence 2, Application US/08938105  
; Patent No. 6353151  
; GENERAL INFORMATION:  
; APPLICANT: Leinwand, Leslie A.  
; APPLICANT: Vaktstrom, Karen L.  
; TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sheridan Ross P.C.  
; STREET: 1700 Lincoln St., Suite 3500  
; CITY: Denver  
; STATE: CO  
; COUNTRY: U.S.A.  
; ZIP: 80203  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/938,105  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Crook, Marnell M.  
; REGISTRATION NUMBER: 31,071  
; REFERENCE/DOCKET NUMBER: 3595-4  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303) 863-9700  
; TELEFAX: (303) 863-0223  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5661 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..5661  
; US-08-938-105-2

Query Match 4.1%; Score 149.4; DB 4; Length 5661;  
Best local similarity 49.2%; Pred. No. 8.2e-31;  
Matches 534; Conservative 0; Mismatches 516; Indels 35; Gaps 4;  
QY 129 TGGAGATTTCACCACTGAGCTGCTTCATTGAGAACTCCGCGGCGGTTCCGGAGGA 188  
|| || ||| |||| || | | |||| | ||| ||

Db 272 TCGTGAACCTTCCTGATGAGCCAGCTGTGCTCTACAAATCTCAAGAGCGCTATGCGGCT 331  
Qy 189 ACCCTATTTATACCTACATCGGTCCTGTCTGTCTCAATCCCTACGAGACCTAC 248  
Db 332 GGATGATCTATACCTACCTACGCTCTGTGTCTGTCACCGCTCAACCCCTATGAAGTGGCTGC 391  
Qy 249 AGATCTACAGCCGCGAGCATATGAAGCCCTACCGTGTGTCTGATTTCTATACATACCAC 308  
Db 392 CAGGTACCAATCGGAGAGTGTAGCTGCTACCGGGGCAAAAGAGCGAGCGCTCCAC 451  
Qy 309 CTCATTTGTTTGCATGGCTGACACTGTATACCGGGCACTTGTACTGAGCGTGGAGC 368  
Db 452 CCCACATCTTCTCCATCTCTGACACAGCTCTATCACTACATCTCTGACAGATCGGAGAAC 511  
Qy 369 AGGCAATGATGATTTCTGAGAGAGTGGGCGAGGACAGACAGAGCCACCAAGAGACTGC 428  
Db 512 AGTCCATCTCTATCACTGTGAGAAATCCGAGCGGGGAGACTGTACACACAGCGTGTCA 571  
Qy 429 TCCAGTTTCTATGC-----AGAGACCTGCCAGCCCTG 461  
Db 572 TCCAGTACTTTGCTAGCATTTGCAACCATAGGGGACCTAGCAAGAGACATCTTAATG 631  
Qy 462 AACGGGTGGCGCATGCGAGACCGCTGTTGAGAGACACCCCGTGTAGAGCGCTTTC 521  
Db 632 CAACCAAGGGCACCTTGAGAGACAGATTTATCCAGGCTAACCTGCTGTGGAGCGCTTTC 691  
Qy 522 GGAATGCCAGACTTCCCGCAAGATTAATCCAGCGGTTTGGAAAGTACATGATGTGC 581  
Db 692 GCAACGCCAAGACTTCCGGAATGACAACTCCCGCTTGGGAGTTTCAATCAGATTC 751  
Qy 582 AGTTTGACTTCAAGGTGCCCCCGTGGAGGCGACATTTCTCAGTTACCTCTGGAAGT 641  
Db 752 ACTTTGAGGCAACGAGAAAGCTGGCTTCTGACACATAGACACTTCTTGGGAGAGT 811  
Qy 642 CCGGGGGTGGCAACAAATCACGAGAGCGGAACCTTCCAGCTCTTTTACAGTACTGG 701  
Db 812 CCGGGTGTATCTTCCAGCTTAAAGGCTGAGAGAACTACCAATTTCTTACAGATTCCT 871  
Qy 702 AGGGGGGAGAGAGACTCTCCGCTGGGCTTGGAAAGGAGAGAGAGAGAGTACT 761  
Db 872 CCAMCAAGAGCGGAGCTGTGACATGTCTGTGTACCAACACCCGCTGACATGATG 931  
Qy 762 TGTACCTGTGAGAGGCGCATGTGCGCAAGTCTCTCCATGACAGACAGAGTACTGGA 821  
Db 932 CTTTGTCTCTAGAGAGAG--GTGTCTGTGGCTCTCATTTGATGATCCGAGAGCTTT 988  
Qy 822 AGCTTATGAGAGAGCGCTGTCCGTCAATGACTTCACTGAGAGTGAAGTGAAGACTTGC 881  
Db 988 TGGCCACTGATAGCCCTTTGATGTGGGGCTTACACAGCAGAGAGAGAGAGCGGTGTCT 1048  
Qy 882 TCAGCATGTGGCGACGCTCTACATCTGGGCAACATCCACTTTGCTGTA---CGAG 937  
Db 1049 ACAAGCTACAGGCGCCATCTATGCAATGAGGAATGAAATTCACAGACAGAGCGGG 1108  
Qy 938 GACAGCAATGCGCAGGTACTACTAGACAGACGCTCAAAATATGACAGAGCTCTTGGT 997  
Db 1109 AGGAGCAGGCGAGCAGAGAGGAGCAAGATGCTGACAAATCTG--CTTACCTCAATGGGG 1167  
Qy 998 GTGGAAGTATCAACACTTAGGAGAGCCCTGACCCACAGAGATCATCCCAAGGGGAA 1057  
Db 1168 CTGAACTACGCGACCTGCTCAAGGGCTGTGTGACCCCTCAGTGAAGTGGGTAAAGAG 1227  
Qy 1058 GAGCTCTGAGCCCACTTGAACCTTGAACAGGCGGCAATATGCAAGGAGTGCCTTCCAG 1117  
Db 1228 TATGTACCAAGGGGCAAGTGTACAGCAGGTGTACTATTTCAATGGGGCACTGGCCAAAG 1287  
Qy 1118 GCTGTATACAGCGGCAATTTACCTGTGTGACAAAAGATCAATAGTCACTGGGCTCT 1177  
Db 1288 TCACTGTACGAGAAAGATGTTCAACTGATGTGTGACACGATCAACGCAACCTTGAGAGC 1347  
Qy 1178 AAGGA 1182  
Db 1348 AAGCA 1352

RESULT 2  
US-09-172-422-2  
Sequence 2, Application US/09172422A  
Patent No. 6300485  
GENERAL INFORMATION:  
APPLICANT: Adams, Arwen E.  
APPLICANT: Chiu, Choi Ying  
APPLICANT: Duhl, David  
APPLICANT: Gorman, Susan W.  
APPLICANT: Leng, Song  
APPLICANT: Sheffield, Val  
APPLICANT: Welch, Juliet  
TITLE OF INVENTION: MYOSIN IXA AND CYCLIC NUCLEOTIDE GATED  
TITLE OF INVENTION: CHANNEL-15 (CNCG-15) POLYNUCLEOTIDES, POLYPEPTIDES,  
FILE REFERENCE: 200130.442  
CURRENT APPLICATION NUMBER: US/09/172,422A  
CURRENT FILING DATE: 1998-10-14  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 8473  
TYPE: DNA  
ORGANISM: Homo sapien  
US-09-172-422-2  
Query Match  
Best Local Similarity 2.9%; Score 104.6; DB 4; Length 8473;  
Matches 288; Conservative 0; Mismatches 284; Indels 3; Gaps 1;  
Qy 143 AGTGAAGCTGCTTCATTTAGAAAGCTCCGCGCGGCTTCCGGAGAGACCTATTTATAC 202  
Db 714 AATGAGAAAACCTCTTTAGAAAACCTAGAGATCGCTTTAGCATGAAAATTTATAC 773  
Qy 203 TACATGCTGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 262  
Db 774 TATGTTGGAGATTTCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 833  
Qy 263 CAGATATGAAGCGTACCGGTGTCTGATTTATGAATGATGATGATGATGATGATGATG 322  
Db 834 AATATATGCTAAATGTATGATTAACCAACCACTGGGAAAACCTGACCAATTTATGCT 893  
Qy 323 GTGGCTGACACTGTATACCGGCACTTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 382  
Db 894 GTGGCTGATGATGATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 953  
Qy 383 TCTGAGAGAGTGGGGCAGGCAAGAGAGGCGCAAGAGAGTGGCTGCTGATTTATGCA 442  
Db 954 TCAGAGAGAGTGGTGTGGGAGAGACTCAAGAGACTCTTTATTTACCACTTACT 1013  
Qy 443 GAGACCTGCCAGCCCTGAAACGGGGTGGCGAGTGGAGACCGCTGTGAGAGCAAC 502  
Db 1014 GCTCTCACTCAAGAA---AGGATTTGCCAGTGTGAGTGAACAGATTTCTGTGAGCTGA 1070  
Qy 503 CCGCTGTGAGGCTTTGGGAATGCCAAGACTCTCCGCAAGCATTAATCTCCAGCGGTTT 562  
Db 1071 CCAGTACTTGGAGGCTTTGGAAATGCAAGAGCTCATTAATTAATTAATTAATTAATTA 1130  
Qy 563 GGAAGTACATGATGTGCGATTTGACTTCAAGAGTGGCCCGTGGAGGCGCAATTTCTC 622  
Db 1131 GGGAGGTTTATTAAGTAAATTAACAGGAACAGAGACTGTACTTGTGCTCATGTGTA 1190  
Qy 623 AGTTACCTCTGAAAAAGTCCCGGTGGTGGCAACCAAAATCACAGGAGAGCACTTCCAC 682  
Db 1191 AAATATCTACTGAGAGAGTCCAGACTCGTTTATTCAGGAGCATTAATGAAGCACTATCAT 1250  
Qy 683 GCTCTTACAGCTACTGAGGAGGCGGAGAGGA 717  
Db 1251 GTATTTATTAATCTCTGTGAGAGAGCAAGTGAAGA 1285



US-08-473-020A-6  
Sequence 6, Application US/08473020A  
Patent No. 5877273  
GENERAL INFORMATION:  
APPLICANT: Hance, Allan J  
APPLICANT: Grandchamp-Desaux, Bernard  
APPLICANT: Levy-Frebault, Veronique  
APPLICANT: Gicquel, Brigitte  
TITLE OF INVENTION: Nucleotide sequences of actinomycetales,  
TITLE OF INVENTION: applications to the synthesis or detection of nucleic  
TITLE OF INVENTION: acids, products of expression of such sequences and  
TITLE OF INVENTION: application as immunogenic compositions.  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Walter H. Dreger  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/473.020A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/623,729  
FILING DATE: 14-DEC-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Walter H  
REGISTRATION NUMBER: 24190  
REFERENCE/DOCKET NUMBER: A54435  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 343 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-473-020A-6

Query Match 1.1%; Score 40.4; DB 2; Length 343;  
Best Local Similarity 53.1%; Pred. No. 0.17; 76; Indels 0; Gaps 0;  
Matches 86; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1267 CGACGACCTTCGATCACTACTGATGAGAACCTGACGACCTTCATCGAGCTGAC 1326  
DB 139 CAACCTCTCTCAAGTTCAGTTCGACATCGAGAGCGGTCGAGAGGTCACCGAGACCT 198  
QY 1327 TCTCAAGTTCGAGCAGAGGATATACGAGGCTGAGGCGATCGCGTGGCAACCTTCACGTA 1386  
DB 199 GCTCAACCGCGCCAGAGAGGTTCGAGACCAAGAGACAGATCGCCGCAACCCCGTGTATCTC 258  
QY 1387 CTTCAACACAGATCATCTGTGACCTGTGAGAGAGAACTT 1428  
DB 259 GGTGGCGCAGCAGTCGATCGGTGACCTGATCGCGCAGAGGCGAT 300

RESULT 6  
US-08-403-852D-4  
Sequence 4, Application US/08403852D  
Patent No. 5891695  
GENERAL INFORMATION:  
APPLICANT: Blanc, Veronique  
APPLICANT: Blanche, Francis  
APPLICANT: Crouzet, Joel

APPLICANT: Jacques, Nathalie  
APPLICANT: Lacroix, Patricia  
APPLICANT: Thibaut, Denis  
APPLICANT: Zagorec, Monique  
APPLICANT: Debussche, Laurent  
APPLICANT: De Crecy-Lagard, Valerie  
TITLE OF INVENTION: Polypeptides involved in the  
TITLE OF INVENTION: biosynthesis of streptogramins, Nucleotide Sequences  
TITLE OF INVENTION: Coding For These Polypeptides And Their Use  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flanagan, Henderson, Farabow, Garrett & Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/403,852D  
FILING DATE: 10-MAY-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR 93/00923  
FILING DATE: 25-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/11441  
FILING DATE: 25-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03806.0054-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1208 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: S.pristinaespiralis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1208  
US-08-403-852D-4

Query Match 1.1%; Score 39.2; DB 2; Length 1208;  
Best Local Similarity 50.0%; Pred. No. 0.68; 98; Indels 0; Gaps 0;  
Matches 98; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 567 ACTATATGATGTGACGTTTACCTTCAAGGTGCCCCCGTGAGGCGACATTTCTCAGTT 626  
DB 347 AGGGGAGAGACACACGCTCGACGACGAGGCGCGGAGACAGGCGCTGATGTGCGCT 406  
QY 627 ACCTCTGGAAGATCCCGGGTGGTGCACCAAAATTCAGAGAGCGGAACCTTCACGCTT 686  
DB 407 AGCGACCGAGACGACCCCTCGCTGATGCCGTGCTGATGACCTTCGACCGCTCTT 466  
QY 687 TTTACCACTACTGAGAGGGGCGAGAGAGACTCTCCGCTGGCGTGGTGAACGGA 746  
DB 467 CGCGCGGCTCACCGAGGTCCGAAAGAGGAGACCGCTCCCTACTCTGCGCCCGAGGGA 526  
QY 747 ACCCCAGAGCTACTT 762  
DB 527 AGACCAAGTCCAT 542

## RESULT 7

US-08-510-646B-4  
Sequence 4, Application US/08510646B  
Patent No. 6077699  
GENERAL INFORMATION:  
APPLICANT: Blanc, Veronique  
APPLICANT: Blanche, Francis  
APPLICANT: Crouzet, Joel  
APPLICANT: Jacques, Nathalie  
APPLICANT: Lacroix, Patricia  
APPLICANT: Thibaut, Denis  
APPLICANT: Zagorec, Monique  
APPLICANT: Debussche, Laurent  
APPLICANT: De Crecy-Lagard, Valerie  
TITLE OF INVENTION: Polypeptides Involved In The  
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
TITLE OF INVENTION: Coding For These Polypeptides And Their Use  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/510,646B  
FILING DATE: 03-AUG-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/403,852  
FILING DATE: 10-MAY-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR 93/00923  
FILING DATE: 25-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/11441  
FILING DATE: 25-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03806.0054-01000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1208 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: S.pristinaespiralis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1208  
US-08-510-646B-4

Query Match 1.1%; Score 39.2; DB 3; Length 1208;  
Best Local Similarity 50.0%; Pred. No. 0.68;  
Matches 98; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

567 AGTACATGATGTGACGTTGACTTCAAGGTCGCCCGCTGGAGGCCACATTCTCAGTT 626

Db 347 AGGCGAGAGACGAGCTGACACAGCGCGCGACGACGAGGCTGATGTTGGCT 406  
QY 627 ACCTCTGGAAGTCCCGGCTGTCACCAAAATCAGGAGCGGAATCTTACAGCT 686  
Db 407 AGCCACCGACGAGACCCCTCGTATGCGCTGCCATCGAGCTGCCACGCTCT 466  
QY 687 TTATCCAGCTACTGAGAGGGGGCGAGAGAGACTCTCCGCTGGGCTTGAACGA 746  
Db 467 CGGGCGGGCTCACCAGAGTCCGCAAGAGGACGCTCCCTACCTGCGCCGAGGCA 526  
QY 747 ACCCCAGAGCTACT 762  
Db 527 AGACCAGGTACCAT 542

## RESULT 8

US-09-231-818-4  
Sequence 4, Application US/09231818  
Patent No. 6171846  
GENERAL INFORMATION:  
APPLICANT: Blanc, Veronique  
APPLICANT: Blanche, Francis  
APPLICANT: Crouzet, Joel  
APPLICANT: Jacques, Nathalie  
APPLICANT: Lacroix, Patricia  
APPLICANT: Thibaut, Denis  
APPLICANT: Zagorec, Monique  
APPLICANT: Debussche, Laurent  
APPLICANT: De Crecy-Lagard, Valerie  
TITLE OF INVENTION: Polypeptides Involved In The  
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
TITLE OF INVENTION: Coding For These Polypeptides And Their Use  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/231,818  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/403,852  
FILING DATE: 10-MAY-1995  
APPLICATION NUMBER: PCT/FR 93/00923  
FILING DATE: 25-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/11441  
FILING DATE: 25-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03806.0054-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1208 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO

ORIGINAL SOURCE:  
ORGANISM: S.pristinaespiralis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1208  
US-09-231-818-4

Query Match  
Best Local Similarity 1.1%; Score 39.2; DB 4; Length 1208;  
Matches 98; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 567 AGTACATGATGTGCGAGTTGCTTCAAGGTGCCCCCGTGGAGCCACATTCAGTT 626  
Db 347 AGGGCGAGGACGACGAGCTCGACGAGGCGCGCGGACGAGGCGCTGATGTTGCGCT 406  
QY 627 ACCTCGTGAAGATCCCGGGTGTGTGCACCAAAATCAGAGGAGGAACTTCACGCT 686  
Db 407 AGCCACCGACGACGAGACCCCTCGCTGATGCCCTGCCATCGAGCTCGCCACCGCTCT 466  
QY 687 TTTACCACTACTGAGGAGGCGGAGAGAGACTCTCGCTGCGCTGGGCTTGAACGGA 746  
Db 467 CGCGCGGCTCAGCGAGGTCCGCAAGAGCGCACCGTCCCTACCTGGCGCCCGACGGA 526  
QY 747 ACCCCAGAGTACTT 762  
Db 527 AGACCCAGTCCACT 542

## RESULT 9

US-08-403-852D-1  
Sequence 1, Application US/08403852D  
Patent No. 5891695  
GENERAL INFORMATION:

APPLICANT: Blanc, Veronique  
APPLICANT: Branche, Francis  
APPLICANT: Crouzet, Joel  
APPLICANT: Jacques, Nathalie  
APPLICANT: Lacroix, Patricia  
APPLICANT: Thibault, Denis  
APPLICANT: Zagorec, Monique  
APPLICANT: Debussche, Laurent  
TITLE OF INVENTION: De Crecy-Lagard, Valerie  
TITLE OF INVENTION: Polypeptides Involved In The  
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/403,852D  
FILING DATE: 10-MAY-1995  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR 93/00923  
FILING DATE: 25-SEP-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: FR 92/11441  
FILING DATE: 25-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03806.0054-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000

TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5392 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: S.pristinaespiralis  
US-08-403-852D-1

Query Match  
Best Local Similarity 1.1%; Score 39.2; DB 2; Length 5392;  
Matches 98; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 567 AGTACATGATGTGCGAGTTGCTTCAAGGTGCCCCCGTGGAGCCACATTCAGTT 626  
Db 3904 AGGGCGAGGACGACGAGCTCGACGAGGCGCGCGGACGAGGCGCTGATGTTGCGCT 3963  
QY 627 ACCTCGTGAAGATCCCGGGTGTGTGCACCAAAATCAGAGGAGGAACTTCACGCT 686  
Db 3964 AGCCACCGACGACGAGACCCCTCGCTGATGCCCTGCCATCGAGCTCGCCACCGCTCT 4023  
QY 687 TTTACCACTACTGAGGAGGCGGAGAGAGACTCTCGCTGCGCTGGGCTTGAACGGA 746  
Db 4024 CGCGCGGCTCAGCGAGGTCCGCAAGAGCGCACCGTCCCTACCTGGCGCCCGACGGA 4083  
QY 747 ACCCCAGAGTACTT 762  
Db 4084 AGACCCAGTCCACT 4099

## RESULT 10

US-08-510-646B-1  
Sequence 1, Application US/08510646B  
Patent No. 6077699  
GENERAL INFORMATION:

APPLICANT: Blanc, Veronique  
APPLICANT: Branche, Francis  
APPLICANT: Crouzet, Joel  
APPLICANT: Jacques, Nathalie  
APPLICANT: Lacroix, Patricia  
APPLICANT: Thibault, Denis  
APPLICANT: Zagorec, Monique  
APPLICANT: Debussche, Laurent  
TITLE OF INVENTION: De Crecy-Lagard, Valerie  
TITLE OF INVENTION: Polypeptides Involved In The  
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/510,646B  
FILING DATE: 03-AUG-1995  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/403,852  
FILING DATE: 10-MAY-1995  
PRIORITY APPLICATION DATA:



APPLICATION NUMBER: PCT/FR 93/00923  
FILING DATE: 25-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/11441  
FILING DATE: 25-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03806.0054-01000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5392 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: S.pristinaespiralis  
US-08-510-646B-1

Query Match 1.1%; Score 39.2; DB 3; Length 5392;  
Best Local Similarity 50.0%; Pred. No. 1.5;  
Matches 98: Conservative 0; Mismatches 98: Indels 0; Gaps 0;

QY 567 AGTACATGATGTGCACTTGTGACTTCAGAGGTCGCCCGGTGGAGGCGCACATTCTCAGTT 626  
DB 3904 AGGGGAGAGACACGACGAGCTGACAGCAGGCGCGGCGCACGACGAGCTGCTGCGCT 3963  
QY 627 ACCTCTGGAAGAGTCCGCGGTGTGACCAAAATCAGCGAGCGGAGCTTCCACGTCT 686  
DB 3964 AGCCACGACGACAGACCCCTGCTGATGCTCCGCTGCGCATCGACCTCGCCACCGCTCT 4023  
QY 687 TTACACGACTACTGAGAGGGGGGCGAGAGAGACTCTCCGTGGCTGGCTTGAACGGA 746  
DB 4024 CGCGCGGCTCACGAGGTCCGCAAGAGGAGCGGACCGTCCCTACCTGCGCGCCCGGACGGCA 4083  
QY 747 ACCCCGAGACTACTT 762  
DB 4084 AGACCCAGGTCAACAT 4099

## RESULT 11

US-09-231-818-1  
Sequence 1, Application US/09231818  
Patent No. 6171846  
GENERAL INFORMATION:  
APPLICANT: Blanc, Veronique  
APPLICANT: Crouzet, Joel  
APPLICANT: Jacques, Nathalie  
APPLICANT: Lacroix, Patricia  
APPLICANT: Thibaut, Denis  
APPLICANT: Zagorec, Monique  
APPLICANT: Debussche, Laurent  
TITLE OF INVENTION: Polypeptides Involved In The  
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
TITLE OF INVENTION: Coding For These Polypeptides And Their Use  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flinnegan, Henderson, Farbow, Garrett & Dunner  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/231,818  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/403,852  
FILING DATE: 10-MAY-1995  
APPLICATION NUMBER: PCT/FR 93/00923  
FILING DATE: 25-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92/11441  
FILING DATE: 25-SEP-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03806.0054-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5392 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: S.pristinaespiralis  
US-09-231-818-1

Query Match 1.1%; Score 39.2; DB 4; Length 5392;  
Best Local Similarity 50.0%; Pred. No. 1.5;  
Matches 98: Conservative 0; Mismatches 98: Indels 0; Gaps 0;

QY 567 AGTACATGATGTGCACTTGTGACTTCAGAGGTCGCCCGGTGGAGGCGCACATTCTCAGTT 626  
DB 3904 AGGGGAGAGACACGACGAGCTGACAGCAGGCGCGGCGCACGAGCTGCTGCGCT 3963  
QY 627 ACCTCTGGAAGAGTCCGCGGTGTGACCAAAATCAGCGAGCGGAGCTTCCACGTCT 686  
DB 3964 AGCCACGACGACAGACCCCTGCTGATGCTCCGCTGCGCATCGACCTCGCCACCGCTCT 4023  
QY 687 TTACACGACTACTGAGAGGGGGGCGAGAGAGACTCTCCGTGGCTGGCTTGAACGGA 746  
DB 4024 CGCGCGGCTCACGAGGTCCGCAAGAGGAGCGGACCGTCCCTACCTGCGCGCCCGGACGGCA 4083  
QY 747 ACCCCGAGACTACTT 762  
DB 4084 AGACCCAGGTCAACAT 4099

## RESULT 12

US-09-103-840A-1/c  
Sequence 1, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
TITLE OF INVENTION: TUBERCULOSIS  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 1  
LENGTH: 441529

```

; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

```

Query Match
Best Local Similarity 1.1%; Score 39.2; DB 4; Length 4411529;
Matches 83; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```

```

QY 1368 CGTGGACACTGTCCAGTCTTCAACAACATCATCTGTGACCTGTGAGAGAGAGT 1427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 641836 CGCTGGAAACCGTCTACTACGGCGCTCGGCGATCAACCCGCTGCGACATGACCGGAGCA 641777
QY 1428 TCAAGGCGCATCTCCATCTTGGATGAAGATGCGCTGCTGGGGAGCGCCAGCAGC 1487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 641776 TCCGCGGCTTCGGCCCAATCTTCCGCCAGTACAGGCGCATCCGAGCGCCCATGTGA 641717
QY 1488 TGACCTTTTGGAGAACTTGGAGAGACACTGTCAAGC 1523
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 641716 TCACCTATCTGCGCCAGGCGCATCAGCAGAGAGC 641681

```

```

RESULT 13
US-09-339-993-1
; Sequence 1, Application US/09339993A
; Patent No. 6040179
; GENERAL INFORMATION:
; APPLICANT: Lex M. Cowser
; TITLE OF INVENTION: ANTISENSE MODULATION OF G-ALPHA-12 EXPRESSION
; FILE REFERENCE: RTS-0064
; CURRENT APPLICATION NUMBER: US/09/339,993A
; CURRENT FILING DATE: 1999-06-25
; NUMBER OF SEQ ID NOS: 47
; SEQ ID NO 1
; LENGTH: 1664
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (78)..(1145)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 1105
; OTHER INFORMATION: unknown
US-09-339-993-1

```

```

Query Match
Best Local Similarity 1.1%; Score 39; DB 3; Length 1664;
Matches 81; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

```

```

QY 106 AGGGGTGACAGACTTTGTCTGTGGAGATTTCCACAGTGAAGCTGCTTCATTGAGAA 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209 AGGGAAGAGCACCATCTCTCAACGATGAAGATCATCCAGAGATGCTTACTCCGAGGA 268
QY 166 CTTCCGCGCGGCTTCGCGGAGAACTCATTTATACCTTCATCGCTCTGTCTAGTCTC 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 269 GGAATGCCGCACTACCGCGGCTTGTCTACAGCAACACCATCCAGTCAATCATGCGCAT 328
QY 226 TGTCAATCCTTACCGAGACCTCAAGATCTAC 256
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 329 TGTCAAGCCATGGCGCACTCAGATCGAC 359

```

```

RESULT 14
US-08-471-033-24
; Sequence 24, Application US/08471033
; Patent No. 5770896
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Koziele, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian

```

```

; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESS: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.308
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,033
; FILING DATE:
; CLASSIFICATION: 530
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: P-40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1399 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..1386
; OTHER INFORMATION: /note="Maize optimized DNA
; OTHER INFORMATION: sequence for VIP2A(a) protein from AB78"
US-08-471-033-24

```

```

Query Match
Best Local Similarity 1.1%; Score 38.6; DB 1; Length 1399;
Matches 174; Conservative 0; Mismatches 199; Indels 6; Gaps 1;

```

```

QY 1190 AGCCCCAGTGGGAGAGCAACCAAGTCTTGGGCTCTCGACATTTACGCTTTGAAGT 1249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 85 AGCATAGCGCTCTGTAACACAGAGGATGATCAAGCGCGAGCTGAAACATCAACAGCCAG 144
QY 1250 TTTACAGATTAACAGCTTCGAGCAGTTCTGCATCACTACTGCATGAGAGAGCTGAGC 1309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 145 AGCAAGTACACCAACCTCAGAACCTGAAGATCAACGACCAAGGTGGAGACTTCAAGAG 204
QY 1310 CTCTTCATCGAGCTGACCTCAAGTCGAGAGAGAGAGATACGAGGCTGAGGCGATCGG 1369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 205 GACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 264
QY 1370 TGGGAAGTGTCCAGTACTTCAACAGAGATCTGTGACCTGGTGAAGAGAGAGAGTTC 1429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 265 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 324
QY 1430 AAGGATCATCTCTCATCTTGTGATGAAGAGTGCCTGCTTGGGAGGCGACGAGACTG 1489

```

Db 325 AAGGAGATCACCCTTCAGCAGCGCCGACGCTTGAG-----GACGAGATCAAGGACCTG 378  
QY 1490 ACCTTTCTGGAGAGTTGGAGCACTGTCAAGCCCACTTCTCTGAGGCAAG 1549  
Db 379 AAGGAGATCAGCAGATGTTCGACAGACCACTGAGCAACGATCATCCTTCAAG 438  
QY 1550 CTCGCTGACCAAGACCA 1568  
Db 439 AACGTGAGCCCAACCA 457

RESULT 15  
US-08-471-044-24  
; Sequence 24, Application US/08471044  
; Patent No. 5840868

GENERAL INFORMATION:  
APPLICANT: Warren, Gregory W  
APPLICANT: Koziel, Michael G  
APPLICANT: Mullins, Martha A  
APPLICANT: Nye, Gordon J  
APPLICANT: Carr, Brian  
APPLICANT: Desai, Nalin M  
APPLICANT: Kostichka, N. Kristy  
APPLICANT: Duck, Nicholas B  
APPLICANT: Estruch, Juan J  
TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains  
NUMBER OF SEQUENCES: 50  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30B

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,044  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/463,483  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40,403  
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8582  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ. ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1399 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..1386

OTHER INFORMATION: /note="Maize optimized DNA  
OTHER INFORMATION: sequence for VIP2A(a) protein from AB78"  
US-08-471-044-24

Query Match 1.1%; Score 38.6; DB 2; Length 1399;  
Best Local Similarity 45.9%; Pred. No. 1.1;  
Matches 174; Conservative 0; Mismatches 199; Indels 6; Gaps 1;

QY 1190 AGCCCGAGCTGGCGGAGCAACACGAGTTCTTGGGCTCCTGAGCATTTACGCTTTGAAGTG 1249  
Db 85 AGCATCAGCCTGCTGTAACAACGAGGTGATCAAGCCGAGCACTGAAATCAACAGCCAG 144  
QY 1250 TTTCAGCATTAACAGCTTCGAGCACTTCGATCACTACTGCAATGAGAGCTGCAAGCAG 1309  
Db 145 AGCAAGTACACCAACCTCCGAACCTGAAGATCACCGACAAAGTGGAGGACTTCAAGAG 204  
QY 1310 CTCTTCATCGAGCTGACTCTCAAGTGGAGCAGAGAGAAATACAGGCTGAGGCAATCGCG 1369  
Db 205 GACCAAGAGAAAGGCCAAGGAGTGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 264  
QY 1370 TGGGAACCTGTCCAGTACTTCACACAAAGATCATCTGTGACCTGGTAGAGAGAGTTG 1429  
Db 265 ACCGAGAGAGGCAAGATGAACAACTTCTGACACAAAGACATCAAGACCACTAC 324  
QY 1430 AAGGAGATCATCTCCATCTTGGATGAAGAGTGCCTGGCTCTGGGAGGCCAGGACCTG 1489  
Db 325 AAGGAGATCACCCTTCAGCATGCGCGGACGCTTCAG-----GACGAGATCAAGGACCTG 378  
QY 1490 ACCTTTCTGGAGAGTTGGAGGACACTGTCAAGCCCACTTCTCTGAGGCAAG 1549  
Db 379 AAGGAGATCGACAGAGATGTTCGACAGACCAACCTGAGCAACGATCATCCTTCAAG 438  
QY 1550 CTCGCTGACCAAGACCA 1568  
Db 439 AACGTGAGCCCAACCA 457

Search completed: December 9, 2002, 18:43:28  
Job time : 12768 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 9, 2002, 14:35:51 ; Search time 109 Seconds  
(without alignments)  
13096.732 Million cell updates/sec

Title: US-09-893-371-3

Perfect score: 3661  
Sequence: 1.ggagcggggcgccggtccg.....ccactgcacclttagcca 3661

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 350425 seqs, 194966369 residues

Total number of hits satisfying chosen parameters: 700850

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PC7\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PC7US\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	342.4	9.4	7465	10	US-09-880-107-3357 Sequence 3357, Ap
2	305.2	8.3	1712	10	US-09-925-300-643 Sequence 643, App
3	299	8.2	3632	12	US-10-044-090-588 Sequence 588, App
4	192.8	5.3	7596	10	US-09-954-456-2215 Sequence 2215, App
5	192	5.2	3048	12	US-10-044-303-3 Sequence 3, App11
6	178	4.9	457	10	US-09-728-445-726 Sequence 726, App
7	174.6	4.8	3501	9	US-09-938-842A-977 Sequence 977, App
8	142.4	3.9	736	10	US-09-925-302-243 Sequence 243, App
9	141.2	3.9	2339	10	US-09-764-853-131 Sequence 131, App
10	118.8	3.2	2345	10	US-09-834-975-1025 Sequence 1025, App
11	104.6	2.9	8473	10	US-09-851-682A-2 Sequence 2, App11
12	98.2	2.7	666	10	US-09-764-853-340 Sequence 340, App
13	95.8	2.6	360	9	US-10-046-935-945 Sequence 945, App
14	95.8	2.6	360	9	US-09-878-178-945 Sequence 945, App
15	93	2.5	476	10	US-09-815-343-787 Sequence 787, App
16	89.8	2.5	376	10	US-09-834-975-638 Sequence 638, App
17	89.8	2.5	376	10	US-09-834-975-664 Sequence 664, App
18	75	2.0	145831	10	US-09-969-708-79 Sequence 79, App1
19	75	2.0	145831	10	US-09-954-456-2116 Sequence 2116, Ap

20	71.6	2.0	558	10	US-09-864-761-6878 Sequence 6878, Ap
21	70.2	1.9	188	10	US-09-864-761-23611 Sequence 23611, A
22	66.4	1.8	574	10	US-09-864-761-14865 Sequence 14865, A
23	65.6	1.8	30013	10	US-09-764-877-3297 Sequence 3297, Ap
24	63.4	1.7	233	10	US-09-864-761-32388 Sequence 32388, A
25	63.4	1.7	590	10	US-09-864-761-15884 Sequence 15884, A
26	61.6	1.7	432	10	US-09-960-352-12994 Sequence 12994, A
27	60.4	1.6	87	10	US-09-864-761-31392 Sequence 31392, A
28	58.8	1.6	491	10	US-09-728-446-1370 Sequence 1370, A
29	52.8	1.4	583	10	US-09-864-761-8935 Sequence 8935, Ap
30	52.2	1.4	2953	10	US-09-866-108-4 Sequence 4, App11
31	52.2	1.4	7707	10	US-09-866-108-2 Sequence 2, App11
32	52.2	1.4	8117	10	US-09-866-108-1 Sequence 1, App11
33	50.2	1.4	90	10	US-09-864-761-25597 Sequence 25597, A
34	45.6	1.2	568	10	US-09-864-761-8095 Sequence 8095, A
35	45.4	1.2	175	10	US-09-864-761-22133 Sequence 22133, A
36	45.4	1.2	307	10	US-09-864-761-24840 Sequence 24840, A
37	45.4	1.2	479	10	US-09-864-761-5359 Sequence 5359, A
38	43.8	1.2	578	10	US-09-864-761-8283 Sequence 8283, Ap
39	43.8	1.2	4701	10	US-09-864-864-326 Sequence 326, App
40	43.2	1.2	302	10	US-09-864-761-23022 Sequence 23022, A
41	43	1.2	311	10	US-09-864-761-17687 Sequence 17687, A
42	43	1.2	311	10	US-09-864-761-25016 Sequence 25016, A
43	41.6	1.1	2451	10	US-09-815-242-9849 Sequence 9849, Ap
44	40.6	1.1	1702	10	US-09-880-107-3680 Sequence 3680, Ap
45	40.2	1.1	3951	9	US-09-712-363-31 Sequence 31, App1

## ALIGNMENTS

RESULT 1  
US-09-880-107-3357  
Sequence 3357, Application US/09880107  
Patent No. US20020142981A1  
GENERAL INFORMATION:  
APPLICANT: Horne, Darci T.  
APPLICANT: Vockley, Joseph G.  
APPLICANT: Scherf, Uwe  
APPLICANT: Gene Logic, Inc.  
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
FILE REFERENCE: 44921-5028-WO  
CURRENT APPLICATION NUMBER: US/09/880,107  
CURRENT FILING DATE: 2001-06-14  
PRIOR APPLICATION NUMBER: US 60/211,379  
PRIOR FILING DATE: 2000-06-14  
PRIOR APPLICATION NUMBER: US 60/237,054  
PRIOR FILING DATE: 2000-10-02  
NUMBER OF SEQ ID NOS: 3950  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3357  
LENGTH: 7465  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. US20020142981A1 U39226  
US-09-880-107-3357  
Query Match 9.4%; Score 342.4; DB 10; Length 7465;  
Best Local Similarity 51.1%; Pred. No. 5,1e-88;  
Matches 1129; Conservative 0; Mismatches 1011; Indels 70; Gaps 11;  
QY 107 GGGGTCCAGGACTTCTCTCGCTGGAGATTTCACAGGAGGCTCCCTTCATGTGGAAC 166  
DB 468 GGGGTGAGGAGCATGTCTCGCTGGGAGGACCTTAC--GAGCGGGCATCTTGGCCAAC 524  
QY 167 CTCGGGCGCGGTTCGGGAGAACCTATTATACCTGCTCTAGTCTCT 226  
DB 525 CTCGTTATCCGCTACCGGAGCACCTCATCTACAGCTATACGGCTCCATCTGGTGGCT 584  
QY 227 GTCATCCCTACCGAGACCTTACAGATCTACAGCGGACCATATGAGACGCTACCGTGT 286  
DB 585 GTGACCCCTACCGAGCTCTCTCATCTACTCGCCAGACGACATCCGCGCAGTATACCAAC 644



```
FILE REFERENCE: PA101
CURRENT APPLICATION NUMBER: US/09/925,300
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05988
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1890
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 643
LENGTH: 1712
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (8)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (1664)
OTHER INFORMATION: n equals a,t,g, or c
US-09-925-300-643
```

Query Match 8.3%; Score 305.2; DB 10; Length 1712;

Best Local Similarity 55.3%; Pred. No. 1.1e-77;

Matches 668; Conservative 2; Mismatches 510; Indels 28; Gaps 3;

```
QY 1046 GCCAAGGGGAGAGGCTGAGCCGACCTGAACTGAAAGCGGCGATATGCAAGGAT 1105
    ||||| ||| ||| ||||| ||| ||||| ||||| ||||| ||||| |||||
Db 265 GCCAAGGAGAGAAATTTCACTACCTGATATGTGGCTCAGGCTTATTTGCCCGAT 324
QY 1106 GCGCTTGCCAGGCTGTATACAGCCGACATTCACCTGGCTGTGCAAAAGATCAATAG 1165
    ||| ||| ||| ||| ||||| ||| ||| ||||| ||| |||||
Db 325 GCTCTGGCTAAAGCTCTACACAGAGTTGTTTCATGTTGTAATCAATCAATG-- 382
QY 1166 TCACCTGGCTTAAGGAGCGCTGAGAGCCGACGCGGGAAGCACCAGCTTCTGGGCTC 1225
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 383 -----AAAGCATTTAAGGCACAAAAGTGGAAGAAAGGATCATGGGTGTT 429
QY 1226 CTGACATTTAGCGCTTTTGAAGTGTTCAGCATTAACAGCTTCGAGAGTTCTGCATCAAC 1285
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 430 CTGACATTTATATGCTTTTGAGATTTTCGAGACACAGCTTTTACGATTTATTTAAT 489
QY 1286 TACTGCAATGAGAAGCTGACAGAGCTTTTCATGAGCTGACATCTCAAGTCGAGAGAG 1345
    ||| ||| ||| ||| ||||| ||| ||| ||||| ||| |||||
Db 490 TATTTGAACAAAGCTGCACAAATCTTCATTTGAATCTCTTAAAGAAAGCAGAGAG 549
QY 1346 GAATAGAGGCTAGGAGCATCGCTGGGAACCTGTCCAGTACTTCAACAACAAGATCATC 1405
    ||| ||| ||| ||| ||||| ||| ||| ||||| ||| |||||
Db 550 GAGTATATACGGAGGATTAAGATGAGACTCACATTTGACTACTTCAATATGCTATCAT 609
QY 1406 TGTGACCTGTAGAGGAGAGATTCAGGGGATCATCTCTTGGATGAAGAGTGCTG 1465
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 610 TGTGACCTATATAGAAATTAACACAAATGGAATCTTGCCCATGCTGATGAAGTGCTC 669
QY 1466 CGTCTGGGAGGCGACGAGCTGACCTTTCTGGAGAAAGTTGGAGACACTGTCAAGCCC 1525
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 670 AGACCTGGACAGTCACTGATGAGACCTTCTTGAAGAAAGCTGAACCAAGATATGCGCAC 729
QY 1526 CACCTCTACTT-----CCGAGCGCAAGAGCTGGCTGACCAAGAGCAGGAATCCCTA 1579
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 730 CACCACTATTTTAAAGCAGAGTGAAGTGTCTCTGCTTCTCAATGACAGCTCTCTG 789
QY 1580 GACCGAGGAGGATTCGCGCTTGTGATTAATGTGAGAGGTGACATGACATGTGACTGGG 1639
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 790 CCGACAGCTGCTTCAGAGATCCAGCATTAATGCTGGAAGAGTGTCTGACAGGTGGAAGA 849
QY 1640 TTTCTGGATTAACCAATGACCTCTCTTCGGAACCTGGAAGAGACATGTCAGCTCA 1699
    ||| ||| ||| ||| ||||| ||| ||| ||||| ||| |||||
Db 850 TTGCTTGACAAAACATGACCTTMTCTATCGAGACCTGTCCCAAGCATGTGGAAGGCC 909
QY 1700 ATGACACCCATCATGCGCCAGTGTCTTTGACAAAGTGAAGTCACTGACAAAGAGCGCCA 1759
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 910 AGCATGCGCTCATCAAGTCTTTGTCCCGGAAGGAAATCCCGCAAGATCAACCTGAAA 969
```

```
QY 1760 GG-----ACGGTGCCACCAGTTCAAGATGAGCCTCTGCGAGCTCGTGAGATCCTG 1812
    ||| ||| ||| ||| ||||| ||| ||| ||||| ||| |||||
Db 970 AGCCCTCCTACACAGAGGCTCAGAGTTCAAGGATCCGTGGCCACTCTGATGAAAAACCTA 1029
QY 1813 AGGCTTAAGAGCCTCCTATATCCGTCATCAAGCCAAAGCAGCCCAAGCAGCGGGT 1872
    ||| ||| ||| ||| ||||| ||| ||| ||||| ||| |||||
Db 1030 CAGACCAAGAACCAACTATATAGTGTATCAAAACCGAATGATAAAAAGCAGCACAC 1089
QY 1873 CGCTTGATGAGTGTCTCATCCGATCAAGGTGAAGTACTGGGACTGATGAGAAATCTG 1932
    ||| ||| ||| ||| ||||| ||| ||| ||||| ||| |||||
Db 1090 ATCTTCAAGAGGCTCTAGTGTCTCATCAATCAGGATCGTGGGCTTTTGGAGAACGTC 1149
QY 1933 CGGCTGCGAGAGCTGCTTGGCTTATTCGCAAAATGAGGCTTTCTCGCAGAGTAC 1992
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1150 CGAGTGGGAGGCGAGGCTACGCTTCAGGAGGCTATGAACCTTGCTAGAAAGATAC 1209
QY 1993 AAGTCACTGTGCCAGAGACATGAGCCATGTGGGAGGAGGAGGAGGAGATGTGGCC 2052
    ||| ||| ||| ||| ||||| ||| ||| ||||| ||| |||||
Db 1210 AAATGCTTTGTAAACAAACATGAGCTCATTTGGAAGAGCAGCAGAGTGTGGAG 1269
QY 2053 GTGTTGTGAGACCTCGGCTCAACGCCAGAGAGTACAAATTTGGCGAGACTAAGATC 2112
    ||| ||| ||| ||| ||||| ||| ||| ||||| ||| |||||
Db 1270 GTCTTATTTAATGAAATTTCCGAGAGAAATCTCTTTGGTATGATCAAGATA 1329
QY 2113 TTGATCCGATTTCCCAAGACCTTATTTGCCACAGAGGACTCCCTGGAAGTCCGGCGAG 2172
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1330 TTGATCCGAAGCCAGAACATTTATTCAAATTTAGAGACCTGAGGAAGCAACCGCTGAG 1389
QY 2173 AGTCTAGCCACCAAGATCCAGCGGCGCTGAGAGGCTTTTCATTTGGCGACAAATTTCTC 2232
    ||| ||| ||| ||| ||||| ||| ||| ||||| ||| |||||
Db 1390 GACTTGGCCACTCTCATTTCAAGATATATCGGGGTGGAATAATGCGCACACTTCTCG 1449
QY 2233 CGGGTGAA 2240
    |||||
Db 1450 CTAATGAA 1457
```

#### RESULT 3

```
US-10-044-090-588
; Sequence 588, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 588
; LENGTH: 3632
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 2131135CB1
US-10-044-090-588
```

Query Match 8.2%; Score 299; DB 12; Length 3632;

Best Local Similarity 56.8%; Pred. No. 1.1e-75;

Matches 596; Conservative 0; Mismatches 440; Indels 13; Gaps 2;

```
QY 1205 AGACACAGGTTCTGGGCTCCGAGCATTTACGAGCTTTGAAGTGTTCAGCATACAGC 1264
    ||| ||| ||| ||| ||||| ||| ||| ||||| ||| |||||
Db 124 AGAAGAAGGTGATGGTGTGTGAGACATTTATGCTTTGAGATTTTCGAGGACAAAGC 183
QY 1265 TTGAGAGAGTTCTGACATCACTGCAATGAGAGAGCTGAGAGCTCTTCATGAGAGCTG 1324
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 184 TTGAGAGAGTTCAATTAATTAATTTAGACGAAAGCTGCAACAAATCTTGAACCTT 243
QY 1325 ACTCTCAAGTCGAGCAGAGGAGGATGAGAGGCTGAGGAGCATCCGCTGGAACCTGTCCAG 1384
    ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
```





Matches	612:	Conservative	0:	Mismatches	600:	Indels	30:	Gaps	3:
QY	137	TTACACAGTGAAGCGTGCCTTCATTTAGAGAACTCCGGCGGGCGGTTCGGGAGAACCTCAATT	196						
Db	298	TACCTCATGTAAACGACGAGTTCCTCCACAATATCCGTGTCTTCAATCAATCAAGATTTAAT	357						
QY	197	TATACGTACATACGGGTCCGTCGCTTGTAGTCCTCTGCATATCCCTACCGGAGACCTTACAGATTCAC	256						
Db	358	TACACCTATTAGGTCTCTTTTGGTTGCCGTCAATTCATTCAGAGAAATTCGAATTTCAC	417						
QY	257	AGCCGGAGACATATGGAAGCGCTACCGTGGTGCAGTTTCTATGAGATACCACTCAATTTCG	316						
Db	418	ACTCAAGAGATGGTTGATATCTTCAAGAGTGTGAGAAAGATGAAAGTTGGCCCACTATTT	477						
QY	317	TTTGCATGGCTGACACTGTATACCGGGCACTGTGTAAGCGTCGGGACCAGGCAAGT	376						
Db	478	TTCCCACTTTTGTATGTTGGCTTATCGTTCAATGTTAGATGATCGTCAAAATCAATCACTC	537						
QY	437	TATCAGAGACCTGCCACGCCCCGTGA-----CGGGGTGGCGCAGTCGAGACCCGCTG	490						
Db	598	CTTGATCTGTGCGCTGGTGCATATCAACCCAAATGATGATGGTATTTGGAACAACAATTT	657						
QY	491	TTTCAGAGCAACCCCGTGTAGAGCCCTTTGGGAATGCCAAGACTCTCCGCAAGATTAAC	550						
Db	658	CTCCAGGACCATCCAAATCCTTGAAGCTTTTGGTAATGCCAAACCCACCGTATCAACAT	717						
QY	551	TTCCAGCCGGTTTGGAAAGTACATGGATGTGCAATTTGACTTCACAAAGGTCCCCCGTGGGA	610						
Db	718	TCATCTCGTTTCGGTAATTCATTTGAATTAATTAATCAACAGTCTGTGTTTATTAGTGT	777						
QY	611	GGCCACATCTTCAGTTTACCTCCTGGAAAAGTCCGGGTGTGACCAAAATCACGGAAG	670						
Db	778	GCTTCATATTCATCTCTACTTTTAGAGAAATCACGTGTCGTTTCCAAATCTGAACCGAA	837						
QY	671	CGGAATCTCCAGCTCTTTTCCAGCTACTGAGGGGGGCGAGGAGAGACTCTCCGTGG	730						
Db	838	CGTAAATATTCACATTTTCTATCAACTCTTACTGTGTGCACCGCGGAAAAGAAAGCT	897						
QY	731	CTGGGCTTGGAAACGAACCCCGACAGCTACTTTTATCTGTGTGAAGGGCAGTGTCCCAAG	790						
Db	898	CTTCACCTTGGCTGGT--CCAGAATCATTTCAACTTAATTAATCAAAAGTGTGTGTGAT	954						
QY	791	GTCCTCTTCATCAACGACAGAGATGACTGGAAGTTATGAGAAAGGCCGTGCCCTCAAT	850						
Db	955	ATCAAAAGGTGTCTGTATGTAGTGAAGAAATTCAAATACACTCGTCAAGCTATGACACTTGT	1014						
QY	851	GACTTCACTGAGATGAAGTGGAGACTTGTCTCAGCATCTGTGGCCAGCGCTCTACATCTG	910						
Db	1015	GTTTCTCTCAAGAAAGAACAAATGTCAATCTTTAAGATCATGTCTGTATCTTACATTTA	1074						
QY	911	GGCAACATCCACTTGTCTGTGACGAGAGACGAATGCCCAGGTTACTACTGAGAACAG	970						
Db	1075	GGTAAACATCAAAATTCGAAAAAGGTGCTGTGAGAGTGTCTTCTCAAAAGACAAAACCGC	1134						
QY	971	CTCAAAATTTGACACAGCGCTCTGGTGTGGAAAGTTACAAACCTTAGGAAGCCCTGACC	1030						
Db	1135	CTCAACGCTGCTTCAACCGCTCTTTGGTGTCAATTCATATGTCCTTGAAGAGGCTCTCATG	1194						
QY	1031	CACAGGAAGATTCATGCCCAAGGGGGAAGAGTCTCTGACGCCCACTGAACCTTTGAACAGCG	1090						
Db	1195	GAAACAGATTTTAAAGCGCGGTGTATTTAGTTGTCAACAATCTCAAGTTCAAAAATTC	1254						
QY	1091	GCATATGCAAGGATCGCTTTGCCAAGGCTGTATACAGCCGCGACATTCACCTGCTGTGC	1150						
Db	1255	TCAATCATCAAGAGACGCTTGTGTCAAAAGCTCTCATATGTGTCTCTTTCTCTCGTTGGTGC	1314						
QY	1151	AGAAAGATCAATAGTCTACTGTGGCTCTTAAGAGCGCTAGAGACCCCAAGCTGGGGAAGCAC	1210						
Db	1315	AAAAAGATCAACATATGTCCTCTGTCAACAGA-----GAAAGCT	1353						

```

RESULT 6
US-09-728-445-726
: Sequence 726, Application US/09728445
: Patent No. US20020102543A1
: GENERAL INFORMATION:
: APPLICANT: Friedrich, Glenn
: APPLICANT: Zambrowicz, Brian
: APPLICANT: Sands, Arthur T.
: TITLE OF INVENTION: No. US20020102543A1e1 Mutated Mammalian Cells and
: TITLE OF INVENTION: Animals
: FILE REFERENCE: LEX-0102-USA
: CURRENT APPLICATION NUMBER: US/09/728,445
: CURRENT FILING DATE: 2000-11-30
: PRIOR APPLICATION NUMBER: US 60/168,358
: PRIOR FILING DATE: 1999-12-01
: NUMBER OF SEQ. ID NOS: 891
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ. ID NO 726
: LENGTH: 457
: TYPE: DNA
: ORGANISM: Mus musculus
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)..(457)
: OTHER INFORMATION: n = A,T,C or G
US-09-728-445-726

```

QY 54 GTGACGGGGTTTCGAGTAGTACCATGG--AGAGGCGCTTGACTGTCGCCGAGACCGGGTAGGGGTG 112  
|||||  
Db 1 GTACAGGTGTTCCAGTAGTACCATTCATTAAGGCGCTTACTGTCGCCGAGACCCNNGTAGGGGTG 60  
QY 113 CAGAGCTTTTGTCCTGCTGTGAGAAATTTTCACCAGTAGGAGCGTGCCTTCATTTAGAACTCCGG 172  
|||||  
Db 61 CAGAGCTTTTGTCCTGCTGTGAGAAATTTTCCAGNAGAGGCTGCCTTCATTTAGAACTCCGCC 120  
QY 173 CGGCGGGTTCCGGGAGAACTCATTTTATACCTACATCGGTCCTGTCTGTCCTG--TGAA 231  
|||||  
Db 121 CGGCGGGTTNNGGAGAACTCATTTATACCTACATCGGTCCTGTGCTGNNCTACTNTNNNNCA 180  
QY 232 TCCCTACCGAGACTTCACATCTACAAGCCGAGACATATGAACTCAACCTGCTGTGAG 291  
|||||  
Db 181 TCCCTACCGAGACTTCACATCTACAGCCCGGCACTTATTTNAAACGGGNNNCGTNAG 240  
QY 292 TTTTCTATGAGTAGACACCTC-----ATTGTGTTGCAGTGGC--TGACACTGTATACCG 342  
|||||  
Db 241 NNAAGTANNNTAANGACNGTMCANCCCTCATTTTNGTTCACAGGCTTGACACTGTATACCG 300  
QY 343 GGCACCTTCTGTAAGTACG--GTGCGGACCAAGCAGTATGATTT-----CTGAGAGAGTG 395  
|||||  
Db 301 GGCACCTTCTGTAAGTACGAGTGCAGNACCAGGACAGTATGATTTTCTGGGNNAGACAGCGGG 360  
QY 396 GGGCGGCAAGACAGAGGCCACACAGA 422  
|||||  
Db 361 GGCAGGCAAGACAGAGGCCACACAGA 387

Query Match	4.8%;	Score 174.6;	DB 9;	Length 3501;
Best Local Similarity	48.6%;	Pred. No. 8.9e-40;		
Matches 656;	Conservative 0;	Mismatches 654;	Indels 39;	Gaps 5;

QY	115	GGACCTTTGCTCGTGGAGAAATTTTCAACAGTAGGCTGCTTCAATTGAGAACTTCGCGG	174
Db	501	GGAGCATCTTATGAGCTTAGTACTTAAATAGAGCCATCAGTGTGTCAACCTCAACTA	560
QY	175	CGCGTCCGGAGAACCTCATTTATACCTACATCGGTCTCGTCCAGTCTGTGCATTC	234
Db	561	TAGTATACACCAAGACATGATATATATACAAAGACAGGCGAGTTTGGTGGCTGGAATTC	620
QY	235	CTACCGAGACCTACATCTACAGCCGGGAGCATATGSAACGCTACCGTGGTGTCAATT	294
Db	621	TTTTAGAGAGTTCCTTTATACGGAATTCGATACATTAACATACAG-----AAGAA	674
QY	295	CTATGAGTACACACCTATTGTTTGGACGTGGCTGACCTGTATACCGGGCACTTGTAC	354
Db	675	ATCAAAATGAGAGTCTCATGTGTATGCAATTGCGAATTCACCAATTCGCGAAATGATAG	734
QY	355	TGAGGCTCGGAGACAGGCGATGATATTTCTGGAGAGATGGGCGACGCAAGACAGGCG	414
Db	735	CGATGAGTAACTCAATCATCATATATATACGGCGGAGAGTGGAGAGGAAATCTAGAC	794
QY	415	CACCAAGAGATGTCCACGTTCTATGCGAGAACCTGCCCCAGCCCTGMAAGGGGTGGCG	474
Db	795	TGCTAAATAGTATGCAATACCTTGGCT-----GCTCTGGAGGGGGAAGCGG	842
QY	475	AGTGGAGACCGCCTGTTGCAAGAACACCCGTGTATGAGGCGCTTTGGGAATGCCAAGC	534
Db	843	GATCGAATATGAGATATCTTAAGACTAATCCCATTTTGGAAAGCATTTGGAAATGCAAAAAC	902
QY	535	TCTCCGCAAGTATACCTCACCGGTTTGGAAAGTACATGAGATGTGCACTTTGACTTCAA	594
Db	903	ATTGGAATATGATTAATCTTACTGTTTGGGAAGCTAAATAGAAATTCATTTTACGGAAG	962
QY	595	GGGTGCCCCGTGGAGGCCACATCTCAGTATACCTCTGTGAAAAAGTCCCGGGTGGTGA	654
Db	963	TGGAAAGATATCGGGTCTCTCAATATCAACCTTTTACTAGAAAAGTCTAAGATGTGTCA	1022
QY	655	CCAAATATCAGSAGAGGGAATCTTCCACGTCCTTTTACCAAGTACTGGAAGGGGCGAGAGA	714
Db	1023	ATGTGCTGAAGGGGAAGGTCAATATCTATTTTATATCAACTTTGTGCTGGGGCTTCAAC	1082
QY	715	GGAGACTCTCCGTGCGCTTGGAGAGGAACCCCAAGAGCTACTTGTACCTGTGGTAA	774



Query Match	3.9%	Score 141.2;	DB 10;	Length 2339;
Best Local Similarity	50.4%;	Pred. No. 3e-30;		
Matches 404;	Conservative 0;	Mismatches 388;	Indels 10;	Gaps 2

QY 1266 TCGACACAGTTCTCGATTCACAACTACTCTCAATGAGAAAGCTCTCAGCAGCTCTTCAATCGAGCTGA 1325  
 Db 131 GGACCACTTTCATTAAGGCGCTGCTGATGTGTATGATTTTGAATCATTTTCTCTGACACACATC 190  
 QY 1266 TCGACACAGTTCTCGATTCACAACTACTCTCAATGAGAAAGCTCTCAGCAGCTCTTCAATCGAGCTGA 1325  
 Db 131 TGGACACAGTTTGGCATTCACAACTAACGCCAATGAGAAAGCTCTCAGCAGCATTTTGTGGCTCACT 250  
 QY 1326 CTCTCAAGTCGGAGCAGAGAGAATACGAGGCGTAGGGGCAATGGCGTGGGAACCTTCCACT 1385  
 Db 251 ACCTAAGGCCCCAGAGAGAGAAATACGAGTTTGAAGGCGCTGAGAGTGGCATTCATCAACT 310  
 QY 1386 ACTTCACACACAGATTCATCTGTGACCTGTGAGAGAGAAATGTAAGGCGATCATCTACCA 1445  
 Db 311 ACCAGGACACACAGCCCTGTTTGGATCTCATTTGAGGGAAGGCCCATACGACTTCGTCTCC 370  
 QY 1446 TCTTGGATGAGAGACTGGCTGGCTGCTGAGGAGGCCACGAGACCTGATCTTGTGAGAGAT 1505  
 Db 371 TCAATAATGAGAAATGCGCGCTCAATAG - - - ACCCAACAGAGGAGGCCAGCTCAGACAC 427  
 QY 1506 TGGAGGACACTGTAAAGCCCCACCCTCACTTCCGACGCAACAACTCGCTGACCAGAGA 1565  
 Db 428 GCATTTAGAGACTGCCCTTGCGAGCGAGCCCTTGCTGGGCGCAAAATMACTCAGCGGGAGC 487  
 QY 1566 CCAGGAATATCCCTAGACCAGAGGGGAGTTCCGCCCTTCTGCAATTATGTGAGAGAGTGACCT 1825  
 Db 488 CCAGCTTCATTTGTGTGCAATTATGCGGGGCGCTGTGGCGTACACACAGCAGCGCTGATG 547  
 QY 1626 ACAGTGTGACTGGGTTTTCTGGATATAAAACAATAGACTCTCTTCCGGAACCTGGAAGAGA 1685  
 Db 548 AGAAGAACAAGAGACCTATCCACCTGAAGCTGAACAGGCTCTCGCAGCAATCCAGAGACC 607  
 QY 1686 CCATGTGCAGCTCAATGAACCCCATCATGGCCCAAGTGTCTTGCACAGAGTGAAGCTAGTG 1745  
 Db 608 CCTCTGCTCATGGGGTGTTTCTCACTAACCCCAAGAAGAACCCAGAGGAACCCCCG 667  
 QY 1746 ACAGAGAAAGGGCC - - - - - AGCAGCGTGGCCACCCAGATTCAAGATGAGACTCTCTCAGC 1798  
 Db 668 GCCAGAGCAAGGGGCCCTGTGTGACCGTGTGTCCAAAGTTCAAGGCGCTCAGCTGAGACAC 727  
 QY 1799 TCGTGTGAGATTCGAGAGTCTAAGAGAGCGCTCCCTATATCCGGTGCATCAAGCCAAACGAGC 1858  
 Db 728 TTCTGAGGTTCTTACACACACACACAGCCGCCCATTAATTCGTGCAATCAAGCCCAACAGCC 787  
 QY 1859 CCAAGCAGCGGGGTGCTTTGATGTAGAGTTCATCCGACATATAGGTGAAGTACCTGGGAC 1918  
 Db 788 AGGGCCAGAGCGGAGACCTTTCTCCAAAGAGAGAGTCTTAGCCAGCTGGAGGCGCTGTGCC 847  
 QY 1919 TCATGGAGAAATTCGGCGCTGGCGCAGAGCTGGCTTTTGCCTATGTCGCAAAATATGAGGCTT 1978  
 Db 848 TCGTGTGAGACCATTCATATCAAGTGTGCTGCGCTTCCCATCGGGTCTCTCAGCAAAACT 907  
 QY 1979 TCTTCGACAGGTGAAGATGAC 2000  
 Db 908 TTGTGAGAACGATACAGTTACT 929

```

? FILE REFERENCE: MRI-016b
? CURRENT APPLICATION NUMBER: US/09/834,975
? CURRENT FILING DATE: 2001-04-13
? PRIOR APPLICATION NUMBER: 60/197,538
? PRIOR FILING DATE: 2000-04-14
? NUMBER OF SEQ ID NOS: 1046
? SOFTWARE: FASTSEQ for Windows Version 4.0.
? SEQ ID NO 1025
? LENGTH: 2345
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (1)...(2345)
? OTHER INFORMATION: n = A,T,C or G
US-09-834-975-1025

```

Query Match	3.2%;	Score 118.8;	DB 10;	Length 2345;
Best Local Similarity	51.0%;	Pred. No. 8.8e-24;		
Matches 422;	Conservative	0;	Mismatches 382;	Indels 24;
				Gaps 5.

QY	563	GGAAAGTACATGATGTGGCAGTTTGAAGCTTCAGAGGTCGCCCGTGGAGGCCACATTTCTC	62
Db	268	GGCAAAATTATTTGCGATCAACACTTTGATGTACTGGCTATATATCTGGGGCCCAATTTGAA	327
QY	623	AGTTACTCCTCGAAAAAGTCCCGGATGTGTGCACCAAAATCACGAGAGCGGAACCTTCCAC	687
Db	328	ACATACCTTTCGAAAAAGTCTCGCTGTTCGTCAAGCCAAAGATGAACTGATTTTCAT	387
QY	683	GTCCTTTACAGGCACTGGAGGGGGGCGAGAGAGACTCTCCGCGCTGGCTTGGAA	744
Db	388	ATCTTTTACAGATTGTT--ATCTGGAGCGAGGAACACCTAAAGCTATTTGGCTTCTTG	445
QY	743	CGAACCCTCCAGACTTCTGTACCTGGAGAGGCCAGTGTGCCAAGCTCTCCATC	802
Db	446	AAGGATTTA-ATACACTACAGGTTTCTCTCCATGGCTATATTCTTA--TTCCGGGACAG	501
QY	803	AAGCACAAAGATCACTGGAAGCTTTATGAGAAAGGCGCTGCCGATTCATTCATCTAGAG	862
Db	502	CAAGCAAAAGATTAATTTCCAGAGAGCCATGTGAAGCAATGACATATATGGGCTTCCAT	561
QY	863	GATGAAGTGGAGAGACTTGTCTGACATTCGAGGGCGACGCTCTACATCTGGCACATCCAC	922
Db	562	GAGAGATTTCTGCAATGCTTTAAAGTAACTTCAGTGTACAGTTTGGAAATTTTCT	621
QY	923	TTTTGCTGTACAGAGACAGCAATGCCAGGTTACTATGAGAACACAGCTCA--AATAT	979
Db	622	TTCAAAAAGAGAAATATCTGATCAACCTTCATGCCAGAAATATACATGTGGCCAGAA	681
QY	980	CTGACACAGGCTCCTGTGTGTGGAAAGTACAACTTAGGGAAGCCCTGACCCACAGAAAG	1037
Db	682	CTCTCCCATCTTCTGGGATGAATGTGATGAGTTTCTCGGGCATTCTGATCCCCGG	741
QY	1040	ATCATTCGCCAAGGGGAGAGACTCCTGAGCCCACTGAACCTTGAAACAGCGGCAATATGCA	1097
Db	742	ATCAAGAGTCCGCCGAGACTATGTCCAAAAAGCCGACCAAAAGACAGGAGATTTTGA	801
QY	1100	AGGATGCGCTTGGCAAGCCTGTACAGCGCGCATTCACCTCGGCTGGGACGAAAAATTC	1157
Db	802	GTAGAGACATTTGGCAAAAGTACTATGAGCGGCTCTTTGGCTGGCTCTTATTCGCATC	861
QY	1160	AATAGTCACTGCGCCTTAGAGACGCTGAGAGCCCAAGCTGGGAGAGACCAAGGTTCT	1219
Db	862	AATTAAGCTTG-----GATAGGACCAAAAGTACAGGAGACATCTTTCAT	906
QY	1220	GGGCTCCGCACTTACGGGCTTTGAAGTCTTACACATTAACAGCTTGCAGCAGTTTCTGC	1279
Db	907	GGATCTCGGATATTTGCTGGATTTGAAATTTTTCAGCTCAACCTCTTTCGAACTTTGC	966
QY	1280	ATCAACTACTGCATGAGAGCTGCAGACGCTCTATGAGCTGCACTGTCAAGTCGAG	1339
Db	967	ATCAACTACCAACATGAGAGCTGACGAGCTGTACACCAACCATGTTTATCTTAAG	1026



; CURRENT FILING DATE: 2002-01-15  
 ; NUMBER OF SEQ ID NOS: 2239  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 945  
 ; LENGTH: 360  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-046-935-945

Query Match  
 Best Local Similarity 55.6%; Score 95.8; DB 9; Length 360;  
 Matches 184; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

QY 1910 ACCGCGACTGATGAGAACTGCGCGAGAGCTGCTTTCCTATGTCGCAAT 1969  
 DB 1 ACCTGGGGCTTTTGGAGAACGTCGAGTGGGAGGCGCTTCAGCGAGCCCT 60  
 QY 1970 ATGAGCCTTTCGCGAGAGTCAAGTCACTGTCGCCAGAGACATGGCCCATGGGCGAG 2029  
 DB 61 ATGACCTTGCTAGAAAGATCAAAATGCTTTGTAACAACATGGCTCATTTGGAAG 120  
 QY 2030 GACGGCCCCAGAGATGCTGTGGCCCTGTTGTCAGACACCTGCGTACAGCCAGAGAGT 2089  
 DB 121 GACCAAGCCAGGCTGTGTGGAGGTCCATTATGATTAATGAATTTCCCGTGAAGAAAT 180  
 QY 2090 ACAAAATGGCGAGGACTAGATCTTCATCCGATTTCCCAAGACCTTATTTGGCCAGAGG 2149  
 DB 181 ACTCCTTTGATGATCAAAAGATATTCAGAAACCAAGACATATTTCAATTAGAG 240  
 QY 2150 ACTGCTTGGAAGTCCGGCGGAGAGTCAAGTCAAGATCCAGGGCGCTGAGGGGCT 2209  
 DB 241 ACCTGAGAGACAGCCGCTGAGAGACTTGGCCACTCTCATTCAGAGATATTCGGGGGT 300  
 QY 2210 TTCAATGGCAGACAGAAATTTCTCCGGGTGA 2240  
 DB 301 GGAATGCCGACACACTTCTGCTAATGAA 331

## RESULT 14

US-09-878-178-945  
 ; Sequence 945, Application US/09878178  
 ; Patent No. US20020177552A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Jiang, Yugu  
 ; APPLICANT: Harlocker, Susan L.  
 ; APPLICANT: Secret, Heather  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
 ; FILE REFERENCE: 210121.527  
 ; CURRENT APPLICATION NUMBER: US/09/878, 178  
 ; CURRENT FILING DATE: 2001-06-08  
 ; NUMBER OF SEQ ID NOS: 2237  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 945  
 ; LENGTH: 360  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-878-178-945

Query Match  
 Best Local Similarity 55.6%; Score 95.8; DB 9; Length 360;  
 Matches 184; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

QY 1910 ACCGCGACTGATGAGAACTGCGCGAGAGCTGCTTTCCTATGTCGCAAT 1969  
 DB 1 ACCTGGGGCTTTTGGAGAACGTCGAGTGGGAGGCGCTTCAGCGAGCCCT 60  
 QY 1970 ATGAGCCTTTCGCGAGAGTCAAGTCACTGTCGCCAGAGACATGGCCCATGGGCGAG 2029  
 DB 61 ATGACCTTGCTAGAAAGATCAAAATGCTTTGTAACAACATGGCTCATTTGGAAG 120  
 QY 2030 GACGGCCCCAGAGATGCTGTGGCCCTGTTGTCAGACACCTGCGTACAGCCAGAGAGT 2089  
 DB 121 GACCAAGCCAGGCTGTGTGGAGGTCCATTATGATTAATGAATTTCCCGTGAAGAAAT 180  
 QY 2090 ACAAAATGGCGAGGACTAGATCTTCATCCGATTTCCCAAGACCTTATTTGGCCAGAGG 2149  
 DB 181 ACTCCTTTGATGATCAAAAGATATTCAGAAACCAAGACATATTTCAATTAGAG 240  
 QY 2150 ACTGCTTGGAAGTCCGGCGGAGAGTCAAGTCAAGATCCAGGGCGCTGAGGGGCT 2209  
 DB 241 ACCTGAGAGACAGCCGCTGAGAGACTTGGCCACTCTCATTCAGAGATATTCGGGGGT 300  
 QY 2210 TTCAATGGCAGACAGAAATTTCTCCGGGTGA 2240  
 DB 301 GGAATGCCGACACACTTCTGCTAATGAA 331

DB 121 GACCAAGCCAGGCTGTGTGGAGGTCTTATTAATGAATTAAGAAATTCCTGAGAGAT 180  
 QY 2090 ACAAAATGGCGAGGACTAGATCTTCATCCGATTTCCCAAGACCTTATTTGGCCAGAGG 2149  
 DB 181 ACTCCTTTGATGATCAAAAGATATTCAGAAACCAAGACATATTCAAATTAAGAG 240  
 QY 2150 ACTGCTTGGAAGTCCGGCGGAGAGTCAAGTCAAGATCCAGGGCGCTGAGGGGCT 2209  
 DB 241 ACCTGAGAGACAGCCGCTGAGAGACTTGGCCACTCTCATTCAGAGATATTCGGGGGT 300  
 QY 2210 TTCAATGGCAGACAGAAATTTCTCCGGGTGA 2240  
 DB 301 GGAATGCCGACACACTTCTGCTAATGAA 331

## RESULT 15

US-09-815-343-787  
 ; Sequence 787, Application US/09815343  
 ; Patent No. US20010053596A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Meagher, Madeleine  
 ; APPLICANT: Xu, Jiangchun  
 ; APPLICANT: King, Gordon E.  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND  
 ; FILE REFERENCE: 210121.504  
 ; CURRENT APPLICATION NUMBER: US/09/815, 343  
 ; CURRENT FILING DATE: 2001-03-22  
 ; NUMBER OF SEQ ID NOS: 1556  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 787  
 ; LENGTH: 476  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc-feature  
 ; LOCATION: (1)..(476)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-815-343-787

Query Match  
 Best Local Similarity 54.8%; Score 93; DB 10; Length 476;  
 Matches 224; Conservative 0; Mismatches 179; Indels 6; Gaps 2;

QY 516 CCTTTGGGAATGCCAAGACTCTCCGACAGATACCTCCAGCGGTTTGAAGATACATG 575  
 DB 1 CTTTCGGGAAAGCGCAAGACCGGTGAAGATGACACTCTCCGCTTGGCAAAATTCATTG 60  
 QY 576 ATGTGACAGTTTGACTCAAGGGTCCCGCTGGAGGCGACATTCAGTACTCTCTGG 635  
 DB 61 GCATCAACTTGTATGTCAATGGCTACATTTGTTGGAGCCACATTTAGACTTATCTTTGG 120  
 QY 636 AAAAGTCCGGGGTGTGACCAAAATCACAGGAGAGGAGGAGTCCAGCTCTTTTACAGG 695  
 DB 121 AGAAATCTGCTGTATCCGGGACGAGCAAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGG 180  
 QY 696 TACTGGAGGGGGGAGAGAGAGACTCTCGTGGCTGGCTTGAAGGAGGAGGAGGAGGAGG 755  
 DB 181 TCTGTGTGGGGCTGAGAGACCTGAAGACCATCTCTGTTGGAGCCGTA---CAACA 237  
 QY 756 GCTACTTTTACTCTGTGTGAAGGCGCATGTGCCAAGGCTCTCTCATCAACGACAAAGAGT 815  
 DB 238 AATACCGCTTCTGTCCAAATGACA---CGTCACATTCCTCCGGGAGCGAGGAGGAGGAGG 294  
 QY 816 ACTGAAGAGTTATAGGAAGGCGCTGCTCATTTACTTCACTAGAGATGAAGTGAAGG 875  
 DB 295 TGTTCACAGAGACCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 924  
 QY 876 ACTGCTTACAGATCTGTGGCGAGGCTCTACATCTGGGAGGAGGAGGAGGAGGAGGAGG 924  
 DB 355 GCGTGTGGGGGATCATCTCAGGGGTTCTTCANCTCGGAGGAGGAGGAGGAGGAGGAGGAGG 403

Tue Dec 10 09:26:51 2002

us-09-893-371-3.rnpb

Page 11

Search completed: December 9, 2002, 16:43:14  
Job time : 145 secs

---







AUTHORS	TITLE	JOURNAL MEDLINE PUBLISHED REFERENCE AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Komori, H., Akiyama, J., Nishikawa, K., Kitazumi, T., Tashiro, H., Itoh, M., Sumi, Y., Ishii, Y., Nakamura, S., Hazama, N., Nishimura, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, M., Okazaki, Y., Ishikawa, T., Ozawa, Y., Tanaka, T., Matsunaga, S., Kawai, J., RIKEN Integrated Sequence Analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)		
Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Saito, T., Okazaki, Y., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, K., Kadota, K., Matsuda, H., Ashburner, M., Batilov, S., Casavant, T., Felsenstein, M., Gaasterland, T., Gissel, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Mikalido, I., Pesole, G., Quackenbush, J., Schiraldi, L.M., Staudt, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okida, T., Furuno, M., Aono, H., Baldarelli, P., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bulc, C., Fletcher, C., Fujita, M., Gariboldi, M., Gusticich, S., Hill, D., Hofmann, M., Hume, D.A., Kamuya, N., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyokawa, O.K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S., and Hayashizaki, Y.		
Functional annotation of a full-length mouse cDNA collection Nature 409 (6821), 685-690 (2001)		
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, P., Bono, H., Brownstein, M., Bulc, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayashi, N., Hill, D., Hiramoto, K., Hirotsugu, T., Horii, F., Kato, H., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koyama, S., Minamizaki, R., Motoyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numata, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schiraldi, L., Shibata, K., Shibata, Y., Shingawa, A., Shitake, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Teijima, Y., Toyota, T., Yamamura, T., Yananaka, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.		
Direct Submission Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Kanagawa 230-0045 Japan (E-mail: genesc.gsc@riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216) Please visit our web site ( <a href="http://genome.gsc.riken.go.jp/">http://genome.gsc.riken.go.jp/</a> ) for further details. CNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer 15' GAGGAGAGAGCGCCGACATCTGAGTTTATTATTTTNNN 3', cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence 15' GAGGAGAGAGCGCCGACATCTGAGTTTATTATTTTAAACCCCCCCC 3'. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.		
AUTHORS	TITLE	JOURNAL MEDLINE PUBLISHED REFERENCE AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Komori, H., Akiyama, J., Nishikawa, K., Kitazumi, T., Tashiro, H., Itoh, M., Sumi, Y., Ishii, Y., Nakamura, S., Hazama, N., Nishimura, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, M., Okazaki, Y., Ishikawa, T., Ozawa, Y., Tanaka, T., Matsunaga, S., Kawai, J., RIKEN Integrated Sequence Analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)		
Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Saito, T., Okazaki, Y., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, K., Kadota, K., Matsuda, H., Ashburner, M., Batilov, S., Casavant, T., Felsenstein, M., Gaasterland, T., Gissel, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Mikalido, I., Pesole, G., Quackenbush, J., Schiraldi, L.M., Staudt, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okida, T., Furuno, M., Aono, H., Baldarelli, P., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bulc, C., Fletcher, C., Fujita, M., Gariboldi, M., Gusticich, S., Hill, D., Hofmann, M., Hume, D.A., Kamuya, N., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyokawa, O.K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S., and Hayashizaki, Y.		
Functional annotation of a full-length mouse cDNA collection Nature 409 (6821), 685-690 (2001)		
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, P., Bono, H., Brownstein, M., Bulc, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayashi, N., Hill, D., Hiramoto, K., Hirotsugu, T., Horii, F., Kato, H., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koyama, S., Minamizaki, R., Motoyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numata, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schiraldi, L., Shibata, K., Shibata, Y., Shingawa, A., Shitake, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Teijima, Y., Toyota, T., Yamamura, T., Yananaka, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.		
Direct Submission Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Kanagawa 230-0045 Japan (E-mail: genesc.gsc@riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216) Please visit our web site ( <a href="http://genome.gsc.riken.go.jp/">http://genome.gsc.riken.go.jp/</a> ) for further details. CNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer 15' GAGGAGAGAGCGCCGACATCTGAGTTTATTATTTTNNN 3', cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence 15' GAGGAGAGAGCGCCGACATCTGAGTTTATTATTTTAAACCCCCCCC 3'. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.		

Db 325 GGGGTGTCAGTTTCTATGAGTAAGTACCACTCATTTGTTGCAAGTGCTGACACTGTATAC 384  
Qy 342 GGGCACTTCTGACTGAGCGCTGGGACCAAGCAAGTATGATTTCTGGAGAGTGGGCGAC 401  
Db 385 GGGCACTTCTGACTGAGCGCTGGGACCAAGCAAGTATGATTTCTGGAGAGTGGGCGAC 444  
Qy 402 GCAAGACAGAGGCCCAAGAGACTGCTCCAGTTCTATGCAAGACCTGGCCAGCCCTG 461  
Db 445 GCAAGACAGAGGCCCAAGAGACTGCTCCAGTTCTATGCAAGACCTGGCCAGCCCTG 504  
Qy 462 AACGGGGGGGGGAGAGAGCGGCTGTTGACAGAGCAACCCCTGTTAGAGGCGCTTG 521  
Db 505 AACGGGGGGGGGAGAGAGCGGCTGTTGACAGAGCAACCCCTGTTAGAGGCGCTTG 564  
Qy 522 GGAATGCAAGACTCTCCGCAACGATTAACCTCCAGCGGTTTGGAAGATGATGATGTC 581  
Db 565 GGAATGCAAGACTCTCCGCAACGATTAACCTCCAGCGGTTTGGAAGATGATGATGTC 624  
Qy 582 AGTTGACTTCAAGGGTGGCCCGGTGGGAGGCGACATTTCTCAGTTACCTCTGGAAGT 641  
Db 625 AGTTGACTTCAAGGGTGGCCCGGTGGGAGGCGACATTTCTCAGTTACCTCTGGAAGT 684  
Qy 642 CCGGGGCTGTGACCAAAATACAGGAGGCGGAACTTCCAGCTTTTACCAAGCTACTG 701  
Db 685 CCGGGGCTGTGACCAAAATACAGGAGAGCGGAACTTCCAGCTTTTACCAAGCTACTG 744  
Qy 702 AGGGGGGAGAGAGAGACTCTCCGCTGGGCTGGGCTTGGAAAGCAACCCAGAGCTACT 761  
Db 745 AGGGGGGAGAGAGAGACTCTCCGCTGGGCTGGGCTTGGAAAGCAACCCAGAGCTACT 804  
Qy 762 TGTACTTGTGAAGGGCCAGTGTGCAAGGTCTCCTCATCAGACAGAGTGTGTA 821  
Db 805 TGTACTTGTGAAGGGCCAGTGTGCAAGGTCTCCTCATCAGAGAGTGTGTA 864  
Qy 822 AGGTTATGAGAGAGCGGCTGTCCTCATTTGACTTCACTGAGATGAAGTGGAGTGC 881  
Db 865 AGGTTATGAGAGAGCGGCTGTCCTCATTTGACTTCACTGAGATGAAGTGGAGTGC 924  
Qy 882 TCAGCATGTGGCCAGCGCTGCTACATCTGGGCAACATCATCTTCTGCTGACGAGCA 941  
Db 925 TCAGCATGTGGCCAGCGCTGCTACATCTGGGCAACATCATCTTCTGCTGACGAGCA 984  
Qy 942 GCAATGCCAGGTTACTACTGAGAACCAAGCTCAAAATATCTGACAGGCTCTTGGTGG 1001  
Db 985 GCAATGCCAGGTTACTACTGAGAACCAAGCTCAAAATATCTGACAGGCTCTTGGTGG 1044  
Qy 1002 AAGGTACAACTATGAGAGCGCTGACCCACAGAGATCATGCGCAAGGGGAGAGAC 1061  
Db 1045 AAGGTACAACTATGAGAGCGCTGACCCACAGAGATCATGCGCAAGGGGAGAGAC 1104  
Qy 1062 TCCTGAGCCCACTGAAGCTTGAACAGGGGCAATATGCAAGGATCGGCTGGCAAGGCTG 1121  
Db 1105 TCCTGAGCCCACTGAAGCTTGAACAGGGGCAATATGCAAGGATCGGCTGGCAAGGCTG 1164  
Qy 1122 TGTACAGCCGCACTTCACTGAGTGTGTCAGAAAGATCAATAGTCACTGGCCTTAAG 1181  
Db 1165 TGTACAGCCGCACTTCACTGAGTGTGTCAGAAAGATCAATAGTCACTGGCCTTAAG 1224  
Qy 1182 ACGGTGAAGAGCCCGAGCTGGGCAAGACACAGGTTCTTGGGCTCCTGACATTTACGCT 1241  
Db 1225 ACGGTGAAGAGCCCGAGCTGGGCAAGACACAGGTTCTTGGGCTCCTGACATTTACGCT 1284  
Qy 1242 TTGAAGTGTTCAGATTAACAGCTTCAGAGCAAGTTCGATCACTACTGACANTGAGAGC 1301  
Db 1285 TTGAAGTGTTCAGATTAACAGCTTCAGAGCAAGTTCGATCACTACTGACANTGAGAGC 1344  
Qy 1302 TGCAGCAGCTTCTTCATGAGCTGACTTCAGCTCGAGAGAGAGAAATGAGAGCTGAG 1361  
Db 1345 TGCAGCAGCTTCTTCATGAGCTGACTTCAGCTCGAGAGAGAGAAATGAGAGCTGAG 1404  
Qy 1362 GCATGCGGTGGAGAACTGTCCAGTACTTCAACACAAAGATCATCTGTGACTGCTGAG 1421  
Db 1405 GCATGCGGTGGAGAACTGTCCAGTACTTCAACACAAAGATCATCTGTGACTGCTGAG 1464

Qy 1422 AGAATTCAGAGGCAATCATCTTCATCTTGGATGAAGAGTGCCTGGCTCCTGGGAGGCA 1481  
Db 1465 AGAATTCAGAGGCAATCATCTTCATCTTGGATGAAGAGTGCCTGGCTCCTGGGAGGCA 1524  
Qy 1482 CGGACCTGACCTTCTTGGAGAAATGAGAGACATGTCGAAGCCCAACCTCTCACTCTGA 1541  
Db 1525 CGGACCTGACCTTCTTGGAGAAATGAGAGACATGTCGAAGCCCAACCTCTCACTCTGA 1584  
Qy 1542 GGCACAACTGCTGACCAAGAGAGCAAGAAATCCCTGAGCCAGGAGGAGTTCGCCCTTC 1601  
Db 1585 GGCACAACTGCTGACCAAGAGAGCAAGAAATCTCTGAGCCAGGAGGAGTTCGCCCTTC 1644  
Qy 1602 TGCATTAATGCTGAGAGAGTGCATCAAGTGTGAGTGGGTTCTGGAATAAAACAATGAC 1661  
Db 1645 TGCATTAATGCTGAGAGAGTGCATCAAGTGTGAGTGGGTTCTGGAATAAAACAATGAC 1704  
Qy 1662 TCCTCTTCCGGAACCTGGAAGAGACCATGTGAGCTGCATGAAACCCCATATGCGCCAGT 1721  
Db 1705 TCCTCTTCCGGAACCTGGAAGAGACCATGTGAGCTGCATGAAACCCCATATGCGCCAGT 1764  
Qy 1722 GCTTGAACAAGAGTGAAGTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1780  
Db 1765 GCTTGAACAAGAGTGAAGTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1824  
Qy 1781 AGATGAGCCTCTGAGAGCTCTGAGAGATCTGAGAGTCTAAGAGGCTGCTATATCCGGT 1840  
Db 1825 AGATGAGCCTCTGAGAGCTCTGAGAGATCTGAGAGTCTAAGAGGCTGCTATATCCGGT 1884  
Qy 1841 GCATCAAGCAAG 1900  
Db 1885 GCATCAAGCAAG 1944  
Qy 1901 AGGTGAAGTACCTGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1960  
Db 1945 AGGTGAAGTACCTGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2004  
Qy 1961 GTCCGAATATGAGGCTTTCCTGAGAGAGTACAGTCACTGTGCCAGAGACATGGCCCA 2020  
Db 2005 GTCCGAATATGAGGCTTTCCTGAGAGAGTACAGTCACTGTGCCAGAGACATGGCCCA 2064  
Qy 2021 TGTGGGAG 2080  
Db 2065 TGTGGGAG 2124  
Qy 2081 CAGAGAGTACAAATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2140  
Db 2125 CAGAGAGTACAAATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2184  
Qy 2141 CCACAG 2200  
Db 2185 CCACAG 2244  
Qy 2201 GGAGGGGCTTTCATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2260  
Db 2245 GGAGGGGCTTTCATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2304  
Qy 2261 AGTCATGAGGAG 2320  
Db 2305 AGTCATGAGGAG 2364  
Qy 2321 AGACATCCGCTGAGTCACTCCGCTTCAATTTTGGGCACTTACCCCGGTCCTGAGA 2380  
Db 2365 AGACATCCGCTGAGTCACTCCGCTTCAATTTTGGGCACTTACCCCGGTCCTGAGA 2424  
Qy 2381 ATGCTTCTTCTTGAAG 2440  
Db 2425 ATGCTTCTTCTTGAAG 2484  
Qy 2441 CCGGGAATGTTTGGAGACACTCTGAGCCACACACCCCAACCTGCTGAGAGAGAGCTGAG 2500  
Db 2485 CCGGGAATGTTTGGAGACACTCTGAGCCACACACCCCAACCTGCTGAGAGAGAGCTGAG 2544

QY	2501	AAC	TGCTACGGGAACTGTGCATGTGAAGAACTATGGTGTGGAAGTACTCCGGAGACTACAGCC	2560
Db	2545	AAC	TCTACGGGAACTGTGCATGTGAAGAACTATGGTGTGGAAGTACTCCGGAGACTACAGCC	2604
QY	2561	CTG	ACTGGACACAGCAGCTGTGAGCAAAAGGCGGTGGCTACTGAAATTTTCAAGGGCANA	2622
Db	2605	CTG	ACTGGACAGAGAGCTGTGAGCAAAAGGCGGTGGCTACTGAAATTTTCAAGGGCANA	2666
QY	2621	AGG	ACAACTACCCCGAGAGTGTCCCGAGACTCTTTCATTTAGCACACGGCTTGGCACAGAGG	2680
Db	2665	AGG	ACAACTACCCCGAGAGTGTCCCGAGACTCTTTCATTTAGCACACGGCTTGGCACAGAGG	2722
QY	2681	AGAT	CAGCCCAAGAGTGTTCATCTCTGGGCTCTGACCACTCAGTATGCCGTGGCCG	2740
Db	2725	AGAT	CAGCCCAAGAGTGTTCATCTCTGGGCTCTGACCACTCAGTATGCCGTGGCCG	2784
QY	2741	TGC	TAAATTTAGACCGTTAAGGCTTACAAGCTGGCCCGGACGCGTGGTGTACAGGCCA	2800
Db	2785	TGC	TAAATTTAGACCGTTAAGGCTTACAAGCTGGCCCGGACGCGTGGTGTACAGGCCA	2844
QY	2801	GTG	CTGTGTGCTATTGTGGAGATGCTTAAAGTCAAGACAGAAATTTGATTATGCCAACTTAA	2866
Db	2845	GTG	CTGTGTGCTATTGTGGAGATGCTTAAAGTCAAGACAGAAATTTGATTATGCCAACTTAA	2904
QY	2861	CCG	GAATCTCTGTACGATGACCTGAGTGAATAGCTTATTTGTCTTCACGTGACGCGTGAAG	2920
Db	2905	CCG	GAATCTCTGTACGATGACCTGAGTGAATAGCTTATTTGTCTTCACGTGACGCGTGAAG	2966
QY	2921	ACA	CAACACAAAGGAGATGTGGTGTGCAAGTATCATGTGTGATCGAGACACTTACCA	2988
Db	2965	ACA	CAACACAAAGGAGATGTGGTGTGCAAGTATCATGTGTGATCGAGACACTTACCA	3024
QY	2981	AGAC	GCGCCCTCAGTGTGACGCGCTGAAACATTTCAACATCAACCGAGGCGACATTAAGT	3040
Db	3025	AGAC	GCGCCCTCAGTGTGACGCGCTGAAACATTTCAACATCAACCGAGGCGACATTAAGT	3084
QY	3041	TTG	CAGGGGATCCAGGACGGGATCATTTGACTTCAATCTGCGGCTCAGAGCTTTCTCA	3100
Db	3085	TTG	CAGGGGATCCAGGACGGGATCATTTGACTTCAATCTGCGGCTCAGAGCTTTCTCA	3144
QY	3101	TTG	ACCAAGGCTTAAATATGGCACTGGCTGTGGTGGCCCGACGGCTGAATTTCTGGGTAT	3166
Db	3145	TTG	ACCAAGGCTTAAATATGGCACTGGCTGTGGTGGCCCGACGGCTGAATTTCTGGGTAT	3204
QY	3161	GAA	GCGCTCGGTGTGACGCGCTCTGTACATCTGATGCTTCCCTTAAGTCCCTCTCCGCTCC	3220
Db	3205	GAA	GCGCTCTCAGTGTGACCGCTCTGTACATCTGATGCTTCCCTTAAGTCCCTCTCCGCTCC	3266
QY	3221	GAC	TTTACCAAAACTCAAGCTTCCAAACAGAGGATCATGAGACCCCTCAAAACCCACGCT	3280
Db	3265	GAG	TTTACCAAAAGCTCAAAAGCTTCCAAACAGAGGATCATGAGACCCCTCAAAACCCACGCT	3322
QY	3281	GCAA	ACTCGGCTTTCTGTGTGCGCCCTCTGTGAGTGTATCAGAGCCAGGAGGCTTACCC	3340
Db	3324	GCAA	ACTCGGCTTTCTGTGTGCGCCCTCTGTGAGTGTATCAGAGCCAGGAGGCTTACCC	3383
QY	3341	ATG	AGTGGGCGGACGCGGCGCACACCAATTTGAANAACAGAGCGTGTAGAGCGCACGGCCA	3400
Db	3384	ATG	AGTGGGCGGACGCGGCGCACACCAATTTGAANAACAGAGCGTGTAGAGCGCACGGCCA	3443
QY	3401	GCC	CTCTGTGATGCCAAATATCTAAGACAAAGGAATTTTAACTGAGGTTTTTCTTGTGAGA	3460
Db	3444	GCC	CTCTGTGATGCCAAATATCTAAGACAAAGGAATTTTAACTGAGGTTTTTCTTGTGAGA	3503
QY	3461	TTTT	TTTTTGTATGCTTTATAGGAACATTTTTTTTAAAGAACCCATTTTTCTAACCTTAAACAC	3520
Db	3504	TTTT	TTTTTGTATGCTTTATAGGAACATTTTTTTTAAAGAACCCATTTTTCTAACCTTAAACAC	3563
QY	3521	ACT	GATGATGTGTTTTTCCCTCGCTCGAACAAGGCGAAGATATGTAACGTAAAGATCTACGCG	3580
Db	3564	ACT	GATGATGTGTTTTTCCCTCGCTCGAACAAGGCGAAGATATGTAACGTAAAGATCTACGCG	3623
QY	3581	GCT	TGGGCTGGAAAGGTTCTCTT--TCTTGGGCAACCCCTTATTCCTTATTCCTGTGCTGCTCTC	3638

	Db	3624	GGTGGCGCTGGAAGGTCCTTCTGTCTGGCAGACCTCTCCATCTGCCTTGCTGTCTGTC	3683
	Oy	3639	CATCCACCGCACACTTTT	3656
	Db	3684	CATCCACCGCACACTTTT	3701
	RESULT 2			
	Bt410257			
	LOCUS			
	DEFINITION	Bt410257	907 bp	mRNA EST 14-AUG-2001
		602966247.1	NCI_CGAP_Lu33	Mus musculus cDNA clone IMAGE:5121610 5'
	ACCESSION			
	VERSION	Bt410257		
	KEYWORDS	Bt410257.1	GI:15171180	
	SOURCE			
	ORGANISM		house mouse.	
			Mus musculus	
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
			Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	
			1 (bases 1 to 907)	
	REFERENCE		NIH-MGC http://mgc.nci.nih.gov/	
	AUTHORS		National Institutes of Health, Mammalian Gene Collection (MGC)	
	TITLE		Unpublished (1993)	
	JOURNAL		Contact: Robert Strausberg, Ph.D.	
	COMMENT		Email: cgapbs-r@mail.nih.gov	
			Tissue Procurement: Gilbert Smith, Ph.D.	
			cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima	
			Bonaldi, Ph.D.	
			cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)	
			DNA Sequencing by: Incyte Genomics, Inc.	
			Clone distribution: NCI-CGAP clone distribution information can be	
			found through the I.M.A.G.E. Consortium/LNLN at:	
			http://image.llnl.gov	
			Plate: LHAM1296 row: n column: 11	
			High quality sequence start: 5	
			High quality sequence stop: 852.	
	FEATURES		Location/Qualifiers	
	SOURCE		1..907	
			/organism="Mus musculus"	
			/strain="CZECH 11"	
			/db_xref="taxon:10090"	
			/clone="IMAGE:5121610"	
			/clone_lib="NCI_CGAP_Lu33"	
			/tissue_type="pooled lung tumors"	
			/lab_host="DH10B (phage-resistant)"	
			/note="Organ: Lung; Vector: pRT3D-Pac (Pharmacia) with a	
			modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st	
			strand cDNA was prepared from mRNA obtained from pooled	
			lung tumors with a Not I - oligo(dT) primer [5'	
			TGTTACCACTCATCTGAAGTGGAGCGGCCGCTCTTTTTTTTTTTT 3'].	
			Double-stranded cDNA was ligated to Eco RI adaptors	
			(Pharmacia), digested with Not I and cloned into the Not	
			I and Eco RI sites of the modified pRT3D vector. Library	
			went through one round of normalization, and was	
			constructed by Bento Soares and M. Fatima Bonaldi. "	
	BASE COUNT		216 a 245 c 267 g 177 t	2 others
	ORIGIN			
	Query Match		22.8%; Score 834.4; DB 13; Length 907;	
	Best Local Similarity		97.7%; Pred. No. 5e-203;	
	Matches	877; Conservative	0; Mismatches 18; Indels 3; Gaps 3;	
	Oy	1473	GGGAGGCCACGGACCTGACCTTCTGGAAGAATTGGAGACACTGTCAAGCCCCACCCTC	1532
	Db	9	GGGAGGCCACGGACCTTCTGGAAGAATTGGAGACACTGTCAAGCCCCACCCTC	68
	Oy	1533	ACTTCCGACGACACAAGTCGCTGACCCAGAAGAACCCAGAAATCCTTAGACCGAGGGAGT	1592
	Db	69	ACTTCCGACGACACAAGTCGCTGACCCAGAAGAACCCAGAAATCCTTAGACCGAGGGAGT	128
	Oy	1593	TTCGCGCTTCGATATATCTGGAGAGTGACCTACAGTGAAGGTTCTCGATATAA	1652

Db 129 TCCGCTTCTGATTATGCTGAGAGTGACCTACAGTGTGACTGGGTTTCTGATATAA 188  
 Oy 1653 ACAATGACCTCTCTTCCGGAAGCTGAAGAGACCATGTCAGCTCAATGAACCCATCA 1712  
 Db 169 ACAATGACCTCTCTTCCGGAAGCTGAAGAGACCATGTCAGCTCAATGAACCCATCA 248  
 Oy 1713 TGGCCAGTCTTTTACACAAGAGTGAAGTCAAGACAGAGGCGCAG -GACGGTGGCCA 1771  
 Db 249 TGGCCAGTCTTTTACACAAGAGTGAAGTCAAGACAGAGGCGCAGAGAGCGTGGCCA 308  
 Oy 1772 CCCATTCAAGATGAGCCCTCTGCAAGCTGTGAGATCTCTGAGGTCTAAGAGAGCTGCT 1831  
 Db 309 CCCATTCAAGATGAGCCCTCTGCAAGCTGTGAGATCTCTGAGGTCTAAGAGAGCTGCT 368  
 Oy 1832 ATATCGGTCATCAAGCAAGCAAGCCAGCCAGCCGCGCTTTGATGAGGTCTCA 1891  
 Db 369 ATATCGGTCATCAAGCAAGCAAGCCAGCCAGCCGCGCTTTGATGAGGTCTCA 428  
 Oy 1892 TCCGACATCAGTGAAGTACCTGGGAGCTGATGAGAGATCTGCGCGTCCGAGAGCTGGCT 1951  
 Db 429 TCCGACATCAGTGAAGTACCTGGGAGCTGATGAGAGATCTGCGCGTCCGAGAGCTGGCT 488  
 Oy 1952 TTGCTATCTGTCGAATATGAGGCTTTCTGCAAGGTACAAAGTCACTGTGCCAGAGA 2011  
 Db 489 TTGCTATCTGTCGAATATGAGGCTTTCTGCAAGGTACAAAGTCACTGTGCCAGAGA 548  
 Oy 2012 CATGGCCCATGTGGGAGAGCGGCCAGAGATGTGTGCGCTGTGTGTCAGACACCTCG 2071  
 Db 549 CATGGCCCATGTGGGAGAGCGGCCAGAGATGTGTGCGCTGTGTGTCAGACACCTCG 608  
 Oy 2072 GCTACAGCCAGAGAAGTACAAATATGAGGAGAGCTAAGATCTTCAATCCGATCCCAAGA 2131  
 Db 609 GCTACAGCCAGAGAAGTACAAATATGAGGAGAGCTAAGATCTTCAATCCGATCCCAAGA 668  
 Oy 2132 CCTTATTTTCCACAGAGAGCTCCCTGGAAGTCCGGCGGAGAGTCTAGCCACCAAGATCC 2191  
 Db 669 CCTTATTTTCCACAGAGAGCTCCCTGGAAGTCCGGCGGAGAGTCTAGCCACCAAGATCC 728  
 Oy 2192 AGCGGCGCTGAGAGGAGCTTCAATTTGGCGAGCAAAATTTCTCGGGGTGAAGCGATCGCA 2251  
 Db 729 AGCGGCGCTGAGAGGAGCTTCAATTTGGCGAGCAAAATTTCTCGGGGTGAAGCGATCGCA 788  
 Oy 2252 TCTGATCCAGTCAATGTCGCTGCGACACTGAGCCGAGAG -AAGCAGCCAAAGAGAG 2310  
 Db 789 TCTGATCCAGTCAATGTCGCTGCGACACTGAGCCGAGAGAGAGCAGCAGCAAGAGAG 848  
 Oy 2311 TGGGAGGCCAGCA -CCATCCGTGCACTCATCCGTGCTTCAATTTCCGCAATTCACACC 2367  
 Db 849 TGGGAGGCCAGCAAGCAATCCGTGCACTCATCCGTGCTTCAATTTCCGCAATTCACACC 906  
  
 RESULT 3 858 bp mRNA linear EST 12-SEP-2001  
 B1653998  
 LOCUS 603280061P1 NCI\_CGAP\_Mam4 Mus musculus cDNA clone IMAGE:5310422 5',  
 DEFINITION mRNA sequence.  
 ACCESSION B1653998  
 VERSION B1653998.1 GI:15568234  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
 NIH-MGC http://mgi.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: c9gabs-remail.nih.gov  
 Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furch  
 Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLML1786 row: 1 column: 15  
 High quality sequence start: 7  
 High quality sequence stop: 775.  
 Location/Qualifiers  
 1. 858  
 /organism="Mus musculus"  
 /strain="NMRI"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5310422"  
 /clone\_1ib="NCI\_CGAP\_Mam4"  
 /tissue.type="tumor, gross tissue"  
 /dev.stage="5 months"  
 /lab\_host="DH10B"  
 /note="Organ: mammary; Vector: pCMV-Sport6; Site\_1: SalI;  
 Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
 Library constructed by Life Technologies. Investigators  
 providing samples: Lothar Hennighausen/Priscilla Furch.  
 NIH Reference for transgenic model: Li et al., Cell Growth  
 and Differentiation 7, 3-11 (1996)."  
 BASE COUNT 211 a 232 c 252 g 163 t  
 ORIGIN  
 Query Match 21.4% Score 783.2; DB 13; Length 858;  
 Best Local Similarity 96.9%; Pred. No. 6.9e-190;  
 Matches 831; Conservative 0; Mismatches 23; Indels 4; Gaps 3;  
  
 Oy 1516 TGTCAAGCCCAACCTCTCTGACGCAAGAGTCTGTGACAGAGAGCAAGAAATC 1575  
 Db 1 TGTCAAGCCCAACCTCTCTGACGCAAGAGTCTGTGACAGAGAGCAAGAAATC 60  
 Oy 1576 CCTAGACGAGGAGATCCGCTTCTGCAATATGCTGAGAGGTACCTACAGTGTAC 1635  
 Db 61 TCTAGACCGAGGAGATCCGCTTCTGCAATATGCTGAGAGGTACCTACAGTGTAC 120  
 Oy 1636 TGGGTTTCTGATATAAACAATGAGACTCTCTTCGGAACCTGGAAGAGCAATGTGAC 1695  
 Db 121 TGGGTTTCTGATATAAACAATGAGACTCTCTTCGGAACCTGGAAGAGCAATGTGAC 180  
 Oy 1696 CTCGAATGAACCCCATCATGAGCCAGTCTTTGACAAGATGAGCTCAGTACAAAGAG 1755  
 Db 181 CTCGAATGAACCCCATCATGAGCCAGTCTTTGACAAGATGAGCTCAGTACAAAGAG 240  
 Oy 1756 GCGAG -GAGGTGTGACCAACCAAGTCAAGATGAGCTCTGCACTGTGAGATCTGAG 1814  
 Db 241 GCGAGAGCGGTGTGACCAACCAAGTCAAGATGAGCTCTGCACTGTGAGATCTGAG 300  
 Oy 1815 GTCTAAGAGCCGCTCTATATCCGCTGTCATCAAGCCAAAGAGCCAGCGGGTGC 1874  
 Db 301 GTCTAAGAGCCGCTCTATATCCGCTGTCATCAAGCCAAAGAGCCAGCGGGTGC 360  
 Oy 1875 CTTTGTATGAGGTGTGTCATCCGACATCAGGTGAAGTACCTGGGAGTATGAGATCTGC 1934  
 Db 361 CTTTGTATGAGGTGTGTCATCCGACATCAGGTGAAGTACCTGGGAGTATGAGATCTGC 420  
 Oy 1935 CGTGGCCAGAGCTGGCTTTGCTTATGTCGCAATATGAGGCTTTCTCTGAGAGTACAA 1994  
 Db 421 CGTGGCCAGAGCTGGCTTTGCTTATGTCGCAATATGAGGCTTTCTCTGAGAGTACAA 480  
 Oy 1995 GTTACGTGTCAGAGACATGAGCCATGTGGGAGAGAGCGCCAGAGATGTGTGCGCGT 2054  
 Db 481 GTTACGTGTCAGAGACATGAGCCATGTGGGAGAGAGCGCCAGAGATGTGTGCGCGT 540  
 Oy 2055 GTTGTGACAGACCTGTGCTACAGGCAAGAGAGTCAAAATATGAGGAGAGTAAATCTT 2114  
 Db 541 GTAGGTGACAGACCTGTGCTACAGGCAAGAGAGTCAAAATATGAGGAGAGTAAATCTT 600  
 Oy 2115 CATCGATTTCCCAAGACCTTATTTGCAACAGAGACTCCCTGGAAGTCCGGCGG -CAGA 2173  
 Db 601 CATCGATTTCCCAAGACCTTATTTGCAACAGAGACTCCCTGGAAGTCCGGCGGCGG -CAGA 660

RESULT 4	
AL583041	
LOCUS	
DEFINITION	939 bp mRNA
AL583041	L7T_NFLO10_BC2 Homo sapiens CDNA clone CSDDL009YJ18 5
* pt1me: mRNA sequence.	
ACCESSION	AL583041
VERSION	AL583041.1 GI:12951623
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 939)  
L4.W.B., Gruber,C., Jassse,J. and Polayer,D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
Contact: Genoscan

1 (bases 1 to 939); Primates; Catarrhini; Homiidae; Homo.  
Eutelestomi;  
Eutelestomi;

BP 191 91006 EVRY cedex - France  
Email: [segre@genoscope.cns.fr](mailto:segre@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Location/Qualifiers  
1..939

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DL009YJ18"
/clone_id="FTI_NF010_0C2"
/sex="male"
/issue_type="B cells from Burkitt lymphoma"
/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with NotI and
cloned into the NotI and EcoRV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : filangellife@life.com URL :
http://fulllength.invitrogen.com"

```

Query Match	20.1%	Score 737.6	DB 9	Length 939
Best Local Similarity	88.0%	Pred. No. 3,7e-178		
Matches 825	Conservative 1	Mismatches 110	Indels 2	Gaps
Qy	1811	TCAGGTTAAGAGACCTCGCTATATCCGGTGATCAAGCCAAAGACGCCAACGACCCG	187	
Db	1	TGCAGCTTAAGAGACCCCTTACGCTCGCGCATCAACCCAAATGATGCCAAACACCCCG	187	
Qy	1871	GTCGCTTTATAGATGCTCATTCGCAATCAGGTGAATACCTGGAGCTGATGAGGAATC	60	
Db	61	GCGGCTTTATACCAAGTCTGATCGCCACAGAGTAAATACCTGGGCTCTTGGAANAAC	120	
Qy	1931	TGCGCGCTGCGAGAGCTGCTTGGCTTACGTGCCAAATATGAGGGCTTTCGCGAGAGCT	1990	

Db	121	TGCGCTGCGC - GA6CCGGCTTTGGCTATTC0CCGCAAAATACGAAGCTTTCTGTCAAAAGG	179
Qy	1991	ACAAGTACACTGTGCCAGAGNATATGGCCATATGTGGGAGAGACGGCCCCACAGATGTGTGG	2051
Db	180	ACAAGTACACTGTGCCAGAGACGTGGCCACGTGGGACAGACGGCCCA - GATGGGGGTGG	238
Qy	2051	CGGTGTGTGTACAGACCTCGGCTACAAAGCCAGAAAGTATCAAAATGGCCAGACTAAGA	2110
Db	239	CTGTCTGTGTGTCCGACACCTGGGCTACAAAGCCAGAAAGTATCAAAAGGGCCAGACCAAG	298
Qy	2111	TCTTCATCCGATTC0CCGCAAAACCTTATTTGGCCACAGAGGATCTCCCGGAAGTCCGGCGGC	2170
Db	299	TCTTCATCCGTTTCC0CCGCAAAACCTGTTTGGCCACAGAGATCCCTGGAGGTCCGGCGGC	358
Qy	2217	AGAGTCTAGCCACCAAGATTCAGGCGGCGCTGGAGGGGCGCTTTCATTTGGCCAGAAATTTTC	2230
Db	359	AGAGCCCTGGGCCCAAAAGATTCAAAGCTGCTGGAGGGGCTTTTCACGTGGGCGCAGAAATTC	418
Qy	2231	TCCGGGGAAGGCAATCAGCATCTGTATTCACATCATGTGTGGCGTGGGCACTGGCCCGGA	2290
Db	419	TCCGGGTGAAGATTCAGCATCTGTATTCACATCATGTGTGGCGTGGGCACTGGCCCGGA	2290
Qy	2291	GGAGGCAAGCAAGAGAGAGTGGGAGCCGACACATTCCTCGATCATCTGGGCTTCA	2350
Db	479	GGAGGCAAGCAAGAGAGAGTGGGAGCCGACACATTCCTCGATCATCTGGGCTTCA	2350
Qy	2351	TTTTCGGCAATTCACCCCGGTGGCCCTTGAAGATGCTTCTTGTGGACCACTGGCGCCT	2410
Db	539	TCTGTGCGCACGGCCCCCGCTGGCCCGCAGAAAGCCTTCTTCTGTGGACCACTGGCGCACCT	598
Qy	2411	CATTTTGTCTTAACCTGAGCGGCAACTGGCCCCGGATTTCTGTGACACTCTGGGCCCA	2470
Db	599	CTTTTGTCTTAACCTGAGCGGCAACTGGCCCCGGATTTCTGTGACACTCTGGGCCCA	2470
Qy	2471	CACCCCAACCTGGCCGTGAGAGAGGCTCTGAAGACTGCTTGGGAACTGTGATGAAGACA	2530
Db	659	CGCCCAACCTGGCCGTGAGAGAGGCTCTGAAGACTGCTTGGGAACTGTGATGAAGACA	2530
Qy	2531	TGTTGTGGAAGTACTGCGGAGACTTACACCCGTGAGTGAAGACAGCTCAGCAAAAGG	2590
Db	719	TGTTGTGGAAGTACTGCGGAGACTTACACCCGTGAGTGAAGACAGCTCAGCAAAAGG	2590
Qy	2591	CGGTGTGTGTAAATTTTCAAGGGCAAGAGACACTACCCCGAGTGTGCCAGAC	2650
Db	779	CGGTGTGTGTAAATTTTCAAGGGCAAGAGACACTACCCCGAGTGTGCCAGAC	2650
Qy	2651	TCTTATTAGACACAGGCTTGGGCAAGAGAGATCAGCCCCAGAGTCTTCAATCTTGG	2710
Db	839	TCTTATTAGACACAGGCTTGGGCAAGAGAGATCAGCCCCAGAGTCTTCAATCTTGG	2710
Qy	2711	GCTCGAACCATTCAGTATGCGGTCGCCGTGCTTAAGA	2748
Db	899	GCTCGAACCATTCAGTATGCGGTCGCTGTGTAA	936

LOCUS B0956845 932 bp mRNA linear EST 21-AUG-2002  
 ACCESSION IMAGE:63939165.5', mRNA sequence.  
 VERSION B0956845  
 KEYWORDS B0956845.1 GI:22372323  
 SOURCE EST.  
 house mouse.  
 Mus musculus.  
 Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 932)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

QY 1327 TCTCAGATCGGAGCAGGAGCAATACGAGGCTGAGGGCATCCGCTGGGACCTGTCCAGTA 1386



Db 61 TCTCAACTGGAGAGAGAGAAATACGAGGCTGAGGACATCGCTGGGAACCTGTCCAGTA 120  
 QY 1387 CTTCAACACAAAGATCATCTGTGACCTGGTAGAGAGAAAGTTCAAGGCATCATCTCCAT 1446  
 Db 121 CTTCAACACAAAGATCATCTGTGACCTGGTAGAGAGAAAGTTCAAGGCATCATCTCCAT 180  
 QY 1447 CTTTGATGAAGAGAGCTGCTGCTGCTGGGAGAGCCAGCACTGACCTTCTGGAGAGATT 1506  
 Db 181 CTTTGATGAAGAGAGCTGCTGCTGCTGGGAGAGCCAGCACTGACCTTCTGGAGAGATT 240  
 QY 1507 GGAGGACACCTGTCACAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1566  
 Db 241 GGAGGACACCTGTCACAGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300  
 QY 1567 CAGGAATATCCCTAGACAGGAGGAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1626  
 Db 301 CAGGAATATCCCTAGACAGGAGGAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360  
 QY 1627 CAGTGTGACTGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1686  
 Db 361 CAGTGTGACTGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420  
 QY 1687 CAGTGTGACTGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1746  
 Db 421 CAGTGTGACTGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480  
 QY 1747 CAGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1805  
 Db 481 CAGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540  
 QY 1806 GATTCCTGAGCTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1865  
 Db 541 GATTCCTGAGCTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600  
 QY 1866 GCGGGGCTGCTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1925  
 Db 601 GCGGGGCTGCTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 660  
 QY 1926 GATTCCTGAGCTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1985  
 Db 661 GATTCCTGAGCTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 720  
 QY 1986 GAGGTACAGTCACTGTGCGCAGAGACATG 2015  
 Db 721 GAGGTACAGTCACTGTGCGCAGAGACATG 750  
 RESULT 7  
 BI65846 760 bp mRNA 1linear EST 12-SEP-2001  
 LOCUS 603289767F1 NCI\_GCAP\_Mam6 Mus musculus cDNA clone IMAGE:5323769 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BI65846  
 VERSION BI65846.1 GI:15580079  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 760)  
 NIH-MGC http://mgc.ncl.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: c9abbs-remail.nih.gov  
 Tissue Procurement: Jeffrey Green M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNLN at:  
 http://image.lnl.gov

Plate: L1821 row: e column: 18  
 High quality sequence stop: 752.  
 Location/Qualifiers  
 1..760  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone\_image="IMAGE:5323769"  
 /clone\_id="NCI\_GCAP\_Mam6"  
 /sex="female, virgin"  
 /tissue\_type="infiltrating ductal carcinoma"  
 /dev\_stage="5 months"  
 /lab\_host="DH10B"  
 /note="Organ: mammary; Vector: pCMV-Sport6; Site:1: SalI;  
 Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
 Library constructed by Life Technologies. Investigator  
 providing samples: Jeffrey Green, M.D., NIH"  
 BASE COUNT 169 a 217 c 224 g 150 t  
 ORIGIN

Query Match 19.5%; Score 714.8; DB 13; Length 760;  
 Best Local Similarity 98.7%; Pred. No. 2.4e-172;  
 Matches 752; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

Db 1714 GGCCCACTGCTTGGACAGAGTGAAGTCAAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1772  
 QY 1773 CCAAGTTCAGATGAGGCTCTGAGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1832  
 Db 61 CCAAGTTCAGATGAGGCTCTGAGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 120  
 QY 1833 TATCCGCTGATCAAGCCCAAG 1892  
 Db 121 TATCCGCTGATCAAGCCCAAG 180  
 QY 1893 CCGACATCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1952  
 Db 181 CCGACATCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 240  
 QY 1953 TGCCTATGCTGCAAAATATGAGGCTTCTGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2012  
 Db 241 TGCCTATGCTGCAAAATATGAGGCTTCTGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 300  
 QY 2013 ATGGCCATGTTGGGAG 2072  
 Db 301 ATGGCCATGTTGGGAG 360  
 QY 2073 CTACAGAGCAGAAAGATCAAAATGAGCAGACTAAGATCTTCAATCCGATTCGCCAAGAC 2132  
 Db 361 CTACAGAGCAGAAAGATCAAAATGAGCAGACTAAGATCTTCAATCCGATTCGCCAAGAC 420  
 QY 2133 CTTATTTGCGACAG 2192  
 Db 421 CTTATTTGCGACAG 480  
 QY 2193 GCGGCGCTGGAAGGAGGCTTCAATTTGGCGACAGAAATTTCTCCGGTGAAGAGAGAGAGAGAG 2252  
 Db 481 GCGGCGCTGGAAGGAGGCTTCAATTTGGCGACAGAAATTTCTCCGGTGAAGAGAGAGAGAGAG 540  
 QY 2253 CTTGATCCAGTCATGTGTGGCTGCGACACTGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 2312  
 Db 541 CTTGATCCAGTCATGTGTGGCTGCGACACTGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
 QY 2313 GCGAGCCAGACATCCGTGACATCCGCGGTTCATTTTGGCGATTCACCCGGGTG 2372  
 Db 601 GCGAGCCAGACATCCGTGACATCCGCGGTTCATTTTGGCGATTCACCCGGGTG 660  
 QY 2373 CCTGAGAGAGCTCTTCTTCTTGGACACAGTGGCGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2432  
 Db 661 CCTGAGAGAGCTCTTCTTCTTGGACACAGTGGCGGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 719  
 QY 2433 GCAACTGCCCGGAGATGTTCTGGACACCTCTGCGCCACAC 2474

```

||||| ||||||| ||||||| |||||||
Db 720 GCAACTG-CGCCGAATGCTCGACACCTCTCGGCCACATC 760

RESULT 8
LOCUS B0942930 935 bp mRNA linear EST 21-AUG-2002
DEFINITION AGNCOURT_8815329 NCI_CGAP_Co24 Mus musculus cDNA clone
IMAGE:6398108 5', mRNA sequence.
ACCESSION B0942930
VERSION B0942930.1 GI:22358408
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 935)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLML1387 row: a column: 21
High quality sequence stop: 617.
Location/Qualifiers
1..935
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="6398108"
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B (TI phage-resistant)"
/note="organ: colon; Vector: pCMV-SPORT6; Site_1: NCI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 222 a 271 c 245 g 197 t
ORIGIN
Query Match 19.5%; Score 714; DB 14; Length 935;
Best Local Similarity 97.8%; Pred. No. 4,3e-172;
Matches 745; Conservative 0; Mismatches 15; Indels 2; Gaps 2;

QY 2826 TAAAGTCAAGCAGAGATGATTATGCGCAACCTAACCGGAATCTCTGTCAGTACCTGAG 2885
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1 TAAAGTCAAGCAGAGATGATTATGCGCAACCTAACCGGAATCTCTGTCAGTACCTGAG 60

QY 2886 TGATAGCCCTATTGTCCTTACGTCGACGCGTGAAGACAACAAGCAGAAGGAGATGTGT 2945
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 61 TGATAGCCCTATTGTCCTTACGTCGACGCGTGAAGACAACAAGCAGAAGGAGATGTGT 120

QY 2946 GCTGCAAGATGATCATGTATGAGACACTAACCAAGAGCGCCCTCAGTGTGACCGCT 3005
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 121 GCTGAGAGATGATCATGTATGAGACACTAACCAAGAGCGCCCTCAGTGTGACCGCT 180

QY 3006 GAACATATCAATCAATCAACGAGCGAGCATTAAGCTTTGAGGGGGTCCAGGCAAGGAGG 3065
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 181 GAACATATCAATCAATCAACGAGCGAGCATTAAGCTTTGAGGGGGTCCAGGCAAGGAGG 240

QY 3066 CATCATTTGACTTACATCGGGCTCAGAGCTTCTCATCACCAAGGCTAAGAAATGGCCACT 3125
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 241 CATCATTTGACTTACATCGGGCTCAGAGCTTCTCATCACCAAGGCTAAGAAATGGCCACT 300

QY 3126 GGCTGTGTGGGCCCAAGGCTGAATTTCTGGTGATGAAGGCTGGGCTGACCGCTCTCA 3185
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 301 GGCTGTGTGGGCCCAAGGCTGAATTTCTGGTGATGAAGGCTGGGCTGACCGCTCTCTCA 360

```

```

QY 3186 CTCCTGATGCTCCCTTCTAGTCCCTCCCTCCCTCCGACTTACCAAAAACCTCAAGCTTCCA 3245
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 361 CTCCTGATGCTCCCTTCTAGTCCCTCCCTCCCTCCGACTTACCAAAAACCTCAAGCTTCCA 420

QY 3246 AACAGGATTCATGACACACCTTCAAAACCCAGCTGCAAACTCTGCTGTGCTGCC 3305
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 421 GACAGGATTCATGACACACCTTCAAAACCCAGCTGCAAACTCTGCTGTGCTGCC 479

QY 3306 CCTCTGAGGTATGACGAGGAGCCAGGAGGCTACCCCAATGATGGGCGGAGCCAGCAC 3365
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 480 CCTCTGAGGTATGACGAGGAGCCAGGAGGCTACCCCAATGATGGGCGGAGCCAGCAC 539

QY 3366 CAATAGAAAAGCAGAGAGCCCTGAGCAGGCGCAGCCCTCTGCTGATGCCAAATATCTA 3425
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 540 CAATAGAAAAGCAGAGAGCCCTGAGCAGGCGCAGCCCTCTGCTGATGCCAAATATCTA 599

QY 3426 AGACAAGGAAATTTTAACGTAGAGTTTCTGTGAGATTTTGTGCTTATAGAAACTA 3485
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 600 AGACAAGGAAATTTTAACGTAGAGTTTCTGTGAGATTTTGTGCTTATAGAAACTA 659

QY 3486 TTTTATTAGAAAAGCATTCTTCTACCTAACACACT-GGATGTCTTTTCCCTGCCCTC 3544
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 660 TTTTATTAGAAAAGCATTCTTCTACCTAACACACTGGAGTGTCTTTTCCCTGCCCTC 719

QY 3545 GAACAGGCGCAGAGATGTAACTGAAGACTGACTGGGCTGG 3586
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 720 GAACAGGCGCAGAGATGTAACTGAAGACTGACTGGGCTGG 761

RESULT 9
LOCUS B1559061 817 bp mRNA linear EST 05-SEP-2001
DEFINITION 603241534F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5294168 5',
mRNA sequence.
ACCESSION B1559061
VERSION B1559061.1 GI:15446375
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 817)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furch
Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLML1744 row: d column: 09
High quality sequence stop: 748.
Location/Qualifiers
1..817
/organism="Mus musculus"
/strain="NMRI"
/db_xref="taxon:10090"
/clone_image="5294168"
/clone_lib="NCI_CGAP_Mam4"
/tissue_type="tumor, gross tissue"
/dev_stage="5 months"
/lab_host="DH10B"
/note="organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Priscilla Furch,
NIH Reference for transgenic model: Li et al., Cell Growth

```

BASE COUNT 196 a 226 c 238 g 157 t (1996)."

## SOURCE

house mouse.

## ORGANISM

Mus musculus

Query Match 19.5%; Score 713.8; DB 13; Length 817;  
Best Local Similarity 95.3%; Pred. No. 4.5e-172;  
Matches 779; Conservative 0; Mismatches 32; Indels 6; Gaps 4;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 763)  
NIH-MGC <http://mgs.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Lohar Hemihausen Ph.D., Chu-Xia Deng Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
<http://image.llnl.gov>  
Plate: LLM1867 row: d column: 17  
High quality sequence stop: 761.  
Location/Qualifiers

FEATURES  
source  
1..763  
/organism="Mus musculus"  
/strain="129,C57BL/6J,FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5341408"  
/clone\_1ib="NCI CGAP Mam3"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="10 months"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-Sport6; Site: 1; Salt:  
Site: 2; NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigators  
providing samples: Lohar Hemihausen/Chu-Xia Deng, NIH  
Reference for transgenic model: Xu et al., Nature Genetics  
22, 37-43 (1999)."

BASE COUNT 188 a 202 c 215 g 157 t 1 others  
ORIGIN

Query Match 19.5%; Score 712.4; DB 13; Length 763;  
Best Local Similarity 98.3%; Pred. No. 1e-171;  
Matches 730; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

QY 1051 GGGGAGAGGCTCTGAGCCGACCTGAACTTGAAGAGGGGCTATGCAAGGATGCGCT 1110  
DB 21 GGGGAGAGGCTCTGAGCCGACCTGAACTTGAAGAGGGGCTATGCAAGGATGCGCT 80

QY 1111 TGCCAAGGCTGTGACGCCGAGCATTACCTGCTGTGAGAAAGATCATATAGTCACT 1170  
DB 81 TGCCAAGGCTGTGACGCCGAGCATTACCTGCTGTGAGAAAGATCATATAGTCACT 140

QY 1171 GGCCTTAAGAGAGCTGAGAGCCCGAGCTGGGAGAGACACACGCTTGGGCTCTGGA 1230  
DB 141 GGCCTTAAGAGAGCTGAGAGCCCGAGCTGGGAGAGACACACGCTTGGGCTCTGGA 200

QY 1231 CATTACGCTTTGAAGTGTTCAGCATTAACAGCTTTCGAGAGTTCCTGCACTACTG 1290  
DB 201 CATTACGCTTTGAAGTGTTCAGCATTAACAGCTTTCGAGAGTTCCTGCACTACTG 260

QY 1291 CAATGAGAGCTGAGACACCTTTCATGAGCTGACTCTCAAGTGGAGAGAGGAATA 1350  
DB 261 CAATGAGAGCTGAGACACCTTTCATGAGCTGACTCTCAAGTGGAGAGAGGAATA 1350

QY 1351 CGAGGCTGAGGAGCTGCGTGGAGACCTGTCCAGTACTTCAACAACAAGATCATCTGGA 1410  
DB 321 CGAGGCTGAGGAGCTGCGTGGAGACCTGTCCAGTACTTCAACAACAAGATCATCTGGA 380

QY 1411 CTTGGTAGAGGAGAGTTCAGAGGCAATCATCTCCATCTGGATGAGAGAGGCTGCTGTC 1470  
DB 381 CTTGGTAGAGGAGAGTTCAGAGGCAATCATCTCCATCTGGATGAGAGAGGCTGCTGTC 440

QY 1471 TGGGAGGCGCAGGAGCTGACCTTTCTGGAGAGTTCGAGAGACTGTGAAGCCGACCC 1530  
DB 441 TGGGAGGCGCAGGAGCTGACCTTTCTGGAGAGTTCGAGAGACTGTGAAGCCGACCC 500

RESULT 10  
LOCUS B1653422 763 bp mRNA EST 12-SEP-2001  
DEFINITION 603301166P1 NCI\_CGAP\_Mam3 Mus musculus cDNA clone IMAGE:5341408 5',  
ACCESSION B1653422  
VERSION B1653422.1 GI:15567658  
KEYWORDS EST.

QY 1531 TCACCTCTGACGACAGCTCGCTGACCAGAGACAGAAATCCCTAGACCGAGGGGA 1590  
 |||||||  
 Db 501 TCACCTCTGACGACAGCTCGCTGACCAGAGACAGAAATCTCTAGACCGAGGGGA 560  
 QY 1531 GTTCCGCTTCTGCAATTAATGCTGAGAGAGTGCCTACAGTGTGAGTGGTCTTCTGATPAA 1650  
 |||||||  
 Db 561 GTTCCGCTTCTGCAATTAATGCTGAGAGAGTGCCTACAGTGTGAGTGGTCTTCTGATPAA 620  
 QY 1651 AAACAAATACCTCTCTCTCCGGAACCTGAAGAGACCATGTGACCTGAATTAACCCCAT 1710  
 |||||||  
 Db 621 AAACAAATACCTCTCTCTCCGGAACCTGAAGAGACCATGTGACCTGAATTAACCCCAT 680  
 QY 1711 CATGGCCAGTGTCTTGTGACAGAGTGTGACAGCAAGAGCCGACGAGCGT-GGC 1769  
 |||||||  
 Db 681 CATGGCCAGTGTCTTGTGACAGAGTGTGACAGCAAGAGCCGACGAGCGTGGGC 740  
 QY 1770 CACCCAGTTCAAGATGAGCCTCC 1792  
 |||||||  
 Db 741 CACCCAGTTCAAGATGAGCCTCC 763

RESULT 11  
 B1156248  
 LOCUS 602903422F1 NCI\_CGAP\_Mam3 Mus musculus cDNA clone IMAGE:5033213 5',  
 DEFINITION mRNA sequence.  
 ACCESSION B1156248  
 VERSION B1156248.1 GI:14616249  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 769)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNLN at:  
 http://image.lnl.gov  
 Plate: L1AM11092 row: k column: 06  
 High quality sequence start: 10  
 High quality sequence stop: 764.

## FEATURES

source  
 1..769  
 /organism="Mus musculus"  
 /strain="129.C57BL/6J,FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5033213"  
 /clone\_lib="NCI\_CGAP\_Mam3"  
 /tissue\_type="tumor, gross tissue"  
 /dev\_stage="10 months"  
 /lab\_host="DH10B"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
 Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
 Library constructed by Life Technologies. Investigators  
 providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH  
 Reference for transgenic model: Xu et al., Nature Genetics  
 22, 37-43 (1999)."

## BASE COUNT

183 a 201 c 226 g 159 t

## ORIGIN

Query Match 19.4%; Score 709.8; DB 13; Length 769;  
 Best Local Similarity 99.5%; Pred. No. 4.7e-171;  
 Matches 733; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 550 CTCACGCCGCTTTGGAAAGTACATGATGTGCAGTTTGACTTCAAGGCTGCCCGCTGGG 609  
 |||||||  
 Db 34 CTCACGCCGCTTTGGAAAGTACATGATGTGCAGTTTGACTTCAAGGCTGCCCGCTGGG 93  
 QY 610 AGGCCAATTCATCTAGTACCTCTCTGAAAAGTCCCGGGTGTGTGCACCAAAATTCAGGAGA 669  
 |||||||  
 Db 94 AGGCCAATTCATCTAGTACCTCTCTGAAAAGTCCCGGGTGTGTGCACCAAAATTCAGGAGA 153  
 QY 670 GCGGAATTCACAGTCTTTTACAGCTACTGGAGGGGGGCGAGAGAGACACTCTCGTGG 729  
 |||||||  
 Db 154 GCGGAATTCACAGTCTTTTACAGCTACTGGAGGGGGGCGAGAGAGACACTCTCGTGG 213  
 QY 730 GCTGGGCTTGGAAAGCAACCCAGAGCTACTTGTACCTGTGTGAAGGGCCAGTGTGCCAA 789  
 |||||||  
 Db 214 GCTGGGCTTGGAAAGCAACCCAGAGCTACTTGTACCTGTGTGAAGGGCCAGTGTGCCAA 273  
 QY 790 GGTCTCTCCATCAACGACAGAGAGTGTGAAGGTTATGAGAGAGCGCTGTCCGTCAAT 849  
 |||||||  
 Db 274 GGTCTCTCCATCAACGACAGAGAGTGTGAAGGTTATGAGAGAGCGCTGTCCGTCAAT 333  
 QY 850 TGACTTCACTGAGAGTGAAGTGAAGAGAGCTTGTGCTACAGCTGTGGCCAGGCTCTACATGT 909  
 |||||||  
 Db 334 TGACTTCACTGAGAGTGAAGTGAAGAGAGCTTGTGCTACAGCTGTGGCCAGGCTCTACATGT 393  
 QY 910 GGGCAACATCCACTTGTGCTGTGACGAGAGACAGCAATGCCAGGTTACTACTGTAGAACA 969  
 |||||||  
 Db 394 GGGCAACATCCACTTGTGCTGTGACGAGAGACAGCAATGCCAGGTTACTACTGTAGAACA 453  
 QY 970 GCTCAATATCTGTGACAGAGCTCTTGTGTGGAAGTTACAACTTAGGAGAGCCTGTGAC 1029  
 |||||||  
 Db 454 GCTCAATATCTGTGACAGAGCTCTTGTGTGGAAGTTACAACTTAGGAGAGCCTGTGAC 513  
 QY 1030 CCACAGGAATCATGCGCAAGGGGGAAGAGCTCTGACGCCCACTGACCTTGAACAGCG 1089  
 |||||||  
 Db 514 CCACAGGAATCATGCGCAAGGGGGAAGAGCTCTGACGCCCACTGACCTTGAACAGCG 573  
 QY 1090 GGCATATGCAAGGAGATGCGCTTGCCAGAGCTGTGTACAGCCGGACA-TTCACCTGGCTGG 1148  
 |||||||  
 Db 574 GGCATATGCAAGGAGATGCGCTTGCCAGAGCTGTGTACAGCCGGACA-TTCACCTGGCTGG 633  
 QY 1149 TCAGAAAGATCAATAGTGTACTGTGCTCTTAAGAGCCTGAGAGCCCTGCGGAAGA 1208  
 |||||||  
 Db 634 TCAGAAAGATCAATAGTGTACTGTGCTCTTAAGAGCCTGAGAGCCCTGCGGAAGA 693  
 QY 1209 CCACGGTCTTGGGCTCCGCAATTTAGCGCTTGAAGTTCAGCAATTAACAGCTGG 1268  
 |||||||  
 Db 694 ACACGGTCTTGGGCTCCGCAATTTAGCGCTTGAAGTTCAGCAATTAACAG-TTCG 752  
 QY 1269 ACACGGTCTTGGGCTCCGCAATTTAGCGCTTGAAGTTCAGCAATTAACAG-TTCG 752  
 |||||||  
 Db 753 ACACGGTCTTGGGCTCCGCAATTTAGCGCTTGAAGTTCAGCAATTAACAG-TTCG 752

## RESULT 12

BF538167

LOCUS 602053638F1 NCI\_CGAP\_Sg2 Mus musculus cDNA clone IMAGE:4193024 5',  
 DEFINITION mRNA sequence.

ACCESSION BF538167

VERSION BF538167.1 GI:11625535

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 861)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LMNL at:  
<http://image.llnl.gov>

Plate: LLM9524 row: k column: 09  
 High quality sequence stop: 748.

## FEATURES

Location/Qualifiers  
 1..861

ORIGIN  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone\_image="4193024"  
 /clone\_lib="NCI\_CGAP\_SG2"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: salivary gland; Vector: pCMVSPORT6; Site: 1;  
 dir: Average insert size 1.3 kb. Constructed by Life  
 Technologies. Note: This is a NCI\_CGAP Library."  
 BASE COUNT 200 a 242 c 242 g 177 t

## Query Match

Best Local Similarity 19.4%; Score 709; DB 12; Length 861;  
 Matches 804; Conservative 0; Mismatches 30; Indels 7; Gaps 7;

Qy 2176 CTAGCCACCAAGATTCAGCGCGCTTGAGGGCTTTTCATTTGCGACAGAAATTTCTCCGG 2235  
 Db 13 CTAGCCACCAAGATTCAGCGCGCTTGAGGGCTTTTCATTTGCGACAGAAATTTCTCCGG 70  
 Qy 2236 GTGAAGCATAGCATCTGTATTCAGTATGTGTGGCTGTGCGACATGGCGCGGAGAG 2295  
 Db 71 GTGAAGCATAGCATCTGTATTCAGTATGTGTGGCTGTGCGACATGGCGCGGAGAG 130  
 Qy 2296 GCACGCCAAGAGAGTGGGAGCCAGACCATCGTGTGACATCGTGTGCTTCAATTTG 2355  
 Db 131 GCACGCCAAGAGAGTGGGAGCCAGACCATCGTGTGACATCGTGTGCTTCAATTTG 190  
 Qy 2356 CGGCATTCACCGCGCTGCTGAGATGCTTCTTTGGACGACGTCGCGCTCATTT 2415  
 Db 191 CGGCATTCACCGCGCTGCTGAGATGCTTCTTTGGACGACGTCGCGCTCATTT 250  
 Qy 2416 TTGCTTAACTGAGAGCGGC-AAGTGGCGGAAATGTTCTGGACACCTCTGGGCC-ACAC 2473  
 Db 251 TTGCTTAACTGAGAGCGGCACACTGCGCGAAATGTTCTGGACACCTCTGGGCCACAC 310  
 Qy 2474 CCCGACCTGCGCTGAGAGAGGCTCAGAACTGCTAGCGGAATGTGATGAAGACATGG 2533  
 Db 311 CCCGACCTGCGCTGAGAGAGGCTCAGAACTGCTAGCGGAATGTGATGAAGACATGG 370  
 Qy 2534 TTGGAAGTGA-CTGCGGAGCATCAGCCCTGAG-TGGAAACGACACTGACGAAAGGC 2591  
 Db 371 TTGGAAGTGA-CTGCGGAGCATCAGCCCTGAG-TGGAAACGACACTGACGAAAGGC 430  
 Qy 2592 GTTGGCTAGTGAATTTTCAAGGGCAAGAAGACACTACCCCAAGAGTGTCCCGACT 2651  
 Db 431 GTTGGCTAGTGAATTTTCAAGGGCAAGAAGACACTACCCCAAGAGTGTCCCGACT 490  
 Qy 2652 CTTCATTAGCACACGGCTTGCGACAGAGAGATCAGCCCCAGAGTCTTCAATCTTGGG 2711  
 Db 491 CTTCATTAGCACACGGCTTGCGACAGAGAGATCAGCCCCAGAGTCTTCAATCTTGGG 550  
 Qy 2712 CTTCGTAACCATTCAGTATGCGCGCTGTGTAATTCGACCGTAAGAGGTGTACAAAGC 2771  
 Db 551 CTTCGTAACCATTCAGTATGCGCGCTGTGTAATTCGACCGTAAGAGGTGTACAAAGC 610  
 Qy 2772 TGGCCCCGGGAGCTGCTGCTAGCGCCAGTGCTGTGTCATTTGGAGAGTGTAAAGT 2831  
 Db 611 TGGCCCCGGGAGCTGCTGCTAGCGCCAGTGCTGTGTCATTTGGAGAGTGTAAAGT 670  
 Qy 2832 CAAGCAGAGAAATGTAATGCCAAGCTAAGCGAATCTCTGTCACTAGCTGAGTGATG 2891  
 Db 671 CAAGCAGAGAAATGTAATGCCA-CTAAGCGAATCTCTGTCACTAGCTGAGTGATG 729

Qy 2892 CCTATTGTGCTTCAGCTGACGCTGAAGACAAACAGACAGAGAGATGTGTGCTGCA 2951  
 Db 730 CTAATTTGTGCTTCAGCTGACGCTGAAGACAAACAGACAGAGATGTGTGCTGCTG 789  
 Qy 2952 GAGGATTCATGTGATTCAGACACTAACCAAGAGCGCTCAGTGTGACCGCTGAACAA 3011  
 Db 790 AAGGATTCCTGTGTGCGGAACCTTAACAAAGAGCTTATGTGTGCGGCTGAACAA 849

Qy 3012 T 3012  
 Db 850 T 850

RESULT 13  
 AL546537  
 LOCUS  
 DEFINITION  
 prime, mRNA sequence.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

Human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 930)  
 Li,W.B., Gruber,C., Jessee,J. and Polayes,D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 CONTACT  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

## FEATURES

Location/Qualifiers

1..930

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CSOD1030YD16"  
 /clone\_lib="LTL\_NFL006\_PL2"  
 /lissue="type="Placenta"  
 /note="Vector: pCMVSPORT 6; Site: 1; NotI; 1st strand cDNA  
 was primed with a NotI-oligo(dT) primer. Five prime end  
 enriched, double-stranded cDNA was digested with Not I and  
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6  
 vector. Library was normalized. Library was constructed by  
 Life Technologies. Contact : Feng Liang Life Technologies,  
 Rockville, Maryland 20850, USA Fax : (1) 301 610 8371  
 Email : fliang@lifetech.com URL :  
<http://fulllength.invitrogen.com>"

BASE COUNT 185 a 278 c 298 g 167 t 2 others

## Query Match

Best Local Similarity 19.1%; Score 700.4; DB 9; Length 930;  
 Matches 769; Conservative 1; Mismatches 97; Indels 1; Gaps 1;

Qy 42 CGGCCCTGGGAGTGAAGAGGCTTGTGATGACATGAGAGCGCTTGACTGCCGAGACC 101  
 Db 64 CGGCCCTGGGAGTGAAGAGGCTTGTGATGACATGAGAGCGCTTGACTGCCGAGACC 123  
 Qy 102 GGGTAGGAGTGAAGAGGCTTGTGATGAGAAATTCACAGAGAGGCTGCTGATG 161  
 Db 124 GGGTAGGAGTGAAGAGGCTTGTGATGAGAAATTCACAGAGAGGCTGCTGATG 183  
 Qy 162 AGAAGCTTCGGGCGGCTTCCGGGAGAACTTATATACCTACATGCTGTCTAG 221  
 Db 184 AGAAGCTTCGGGCGGCTTCCGGGAGAACTTATATACCTACATGCTGTCTAG 243  
 Qy 222 TCTGTGCAATTCCTACCGAGCTTACAGATCTAGAGCGGCGAGCTATAGGAAGCTTACC 281  
 Db 244 TCTGTGCAATTCCTACCGAGCTTACAGATCTAGAGCGGCGAGCTATAGGAAGCTTACC 303

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	JOURNAL	COMMENT
282	GTGGGTGATAGTTTCTATGAATGACACCTCATATTGTGGAGTGGCTGACATGTAATAC	BM450184	5	NIH-MGC	human.	Homo sapiens	1 (bases 1 to 1031)	NIH-MGC	http://mhc.nci.nih.gov/	
304	GTGGGTGACGTTCTTATGAATGACCCCTTACACTCTTTTCCGTGGCGGACACTGTGTACC	BM450184	5	NIH-MGC	human.	Homo sapiens	1 (bases 1 to 1031)	NIH-MGC	http://mhc.nci.nih.gov/	
342	GGGCACTTCGTAAGCGTGGGACCAAGGACAGTATGATTTCTGAGAGAGATGGGGCAG	BM450184	5	NIH-MGC	human.	Homo sapiens	1 (bases 1 to 1031)	NIH-MGC	http://mhc.nci.nih.gov/	
364	GAGCACTCGCAACGGAGCGTGGGACCAAGGCTGTGATGATCTCTTGGGAGACGGGGCAG	BM450184	5	NIH-MGC	human.	Homo sapiens	1 (bases 1 to 1031)	NIH-MGC	http://mhc.nci.nih.gov/	
402	GCAGACAGAGAGCCCAAGAGACTGCTCCAGTTTATATGACAGACCTTGGCCAGCCCTTG	BM450184	5	NIH-MGC	human.	Homo sapiens	1 (bases 1 to 1031)	NIH-MGC	http://mhc.nci.nih.gov/	
424	GCAGACAGAGAGCCCAAGAGAGCTGCTGCAAGTTTATATGACAGACCTTGGCCAGCCCG	BM450184	5	NIH-MGC	human.	Homo sapiens	1 (bases 1 to 1031)	NIH-MGC	http://mhc.nci.nih.gov/	
462	AACGGGTGGCGCATGTCGGAGACCCCTGTTGGACAGACACCCCTGTTAGAGGCTTTTG	BM450184	5	NIH-MGC	human.	Homo sapiens	1 (bases 1 to 1031)	NIH-MGC	http://mhc.nci.nih.gov/	
484	AGCGGGAGAGTGTCCCTGGGGGACCGGCTGTACAGACGACACCGGCTGCTGGAGGCTTTG	BM450184	5	NIH-MGC	human.	Homo sapiens	1 (bases 1 to 1031)	NIH-MGC	http://mhc.nci.nih.gov/	
522	GGATGCCAAGACTCTCCGCAACGATTAACCTCCAGCGGTTTGGAAAGTACATGATGTTC	BM450184	5	NIH-MGC	human.	Homo sapiens	1 (bases 1 to 1031)	NIH-MGC	http://mhc.nci.nih.gov/	
544	GAATTCGCCAAGACCTTCGGAACGATTAACCTCCAGGATGCGGAAGTACATGATGTTC	BM450184	5	NIH-MGC	human.	Homo sapiens	1 (bases 1 to 1031)	NIH-MGC	http://mhc.nci.nih.gov/	
582	AGTTGACTTCAAGGTCCTCCCGTGGGAGGCGCAATTCGTGATTAAGTCTCTGGAAAGT	BM450184	5	NIH-MGC	human.	Homo sapiens	1 (bases 1 to 1031)	NIH-MGC	http://mhc.nci.nih.gov/	
604	AGTTGACTTCAAGGTCCTCCCGTGGGAGGCGCAATTCGTGATTAAGTCTCTGGAAAGT	BM450184	5	NIH-MGC	human.	Homo sapiens	1 (bases 1 to 1031)	NIH-MGC	http://mhc.nci.nih.gov/	
642	CCCGGTCGTCACCAAAATCAACGAGAGCGCAATTCGACGTCCTTTTACGACGTAACG	BM450184	5	NIH-MGC	human.	Homo sapiens	1 (bases 1 to 1031)	NIH-MGC	http://mhc.nci.nih.gov/	
663	CACGAGTGTGTCACCAAGATCATGTGGGAGCGCAATTCGACGTCCTTTTACGACGTAACG	BM450184	5	NIH-MGC	human.	Homo sapiens	1 (bases 1 to 1031)	NIH-MGC	http://mhc.nci.nih.gov/	
702	AGGGGGCGAGAGAGACTCTCCGTCGCTGGGCTTGGAAAGCAACCCCGACGACTACT	BM450184	5	NIH-MGC	human.	Homo sapiens	1 (bases 1 to 1031)	NIH-MGC	http://mhc.nci.nih.gov/	
723	AGGGGGCGAGAGAGACTCTTCGACGAGTGTGGCTTGGAAAGCAACCCCGACGACTACT	BM450184	5	NIH-MGC	human.	Homo sapiens	1 (bases 1 to 1031)	NIH-MGC	http://mhc.nci.nih.gov/	
762	TGTACTCGTGAAGGGCCAGTGTGGCAAGTCTCTCATCAACGCAAGAGTACTGGA	BM450184	5	NIH-MGC	human.	Homo sapiens	1 (bases 1 to 1031)	NIH-MGC	http://mhc.nci.nih.gov/	
783	TGTACTCGTGAAGGGCCAGTGTGGCAAGTCTCTCATCAACGCAAGAGTACTGGA	BM450184	5	NIH-MGC	human.	Homo sapiens	1 (bases 1 to 1031)	NIH-MGC	http://mhc.nci.nih.gov/	
822	AGGTTATGAGAGAGCGCTGTCCTCATTTGACTTCACTGAGATGAAGTGGAGACTTGC	BM450184	5	NIH-MGC	human.	Homo sapiens	1 (bases 1 to 1031)	NIH-MGC	http://mhc.nci.nih.gov/	
843	AAGTCGTAAGGAGGCTCTGACAGTCATTTGATTTACCGAGATGAAGTGGAGGACTTGC	BM450184	5	NIH-MGC	human.	Homo sapiens	1 (bases 1 to 1031)	NIH-MGC	http://mhc.nci.nih.gov/	
882	TCAGCATCTGTGGCCAGCGTCTCAATCT	BM450184	5	NIH-MGC	human.	Homo sapiens	1 (bases 1 to 1031)	NIH-MGC	http://mhc.nci.nih.gov/	
903	TGAGCATCTGTGGCCAGCGTCTCAATCT	BM450184	5	NIH-MGC	human.	Homo sapiens	1 (bases 1 to 1031)	NIH-MGC	http://mhc.nci.nih.gov/	

FEATURES		Location/Qualifiers	
SOURCE		1. 1031	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="IMAGE:5528227"	
		/clone_lib="NH_MGC_72"	
		/tissue_type="melanotic melanoma"	
		/lab_host="DH10B (phage-resistant)"	
		/note="Organ: Skin; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo Average insert size 2 kb. Library constructed by Life Technologies."	
BASE COUNT	229 a 298 c 318 g 186 t		
ORIGIN			
Query Match	18.6%; Score 679.8; DB 13; Length 1031;		
Best Local Similarity	85.9%; Pred. No. 2,7e-163;		
Matches 801; Conservative	0; Mismatches 127; Indels 5; Gaps		
Qy	1405	CTGTGACCTGGTAGAGAGAGATTCAAGGGCATCTCTCCATCTTGTGATGTAAGAGTGGCT	1464
Db	1	CTGTGATCTGTGTGAGAGAGAAATTAAAGGCATCACTCTGATTTTGTATGTAGAGAGTGTCT	60
Qy	1465	GGCTGCTGGGGAGGCGACGCACTGTGACTTTCGGAGAGTTGGAGAGACAGTGTCAACCC	1524
Db	61	GGCGCCGGGAGGCGACACACCTGACTTCTCGGAGAAAGCTGGAGAGTACTGTCAACGA	120
Qy	1525	CCACCTCTACCTTCTGTGACGACCAAGCTCGCTGACACAGAGACACAGAAATCCTTAGACG	1584
Db	121	CCATTCACACTCTCTGTAGCGACAAAGCTGGCTGACACAGCGGACAGGAATCTCTGGCGC	180
Qy	1585	AGGGGAGTCCCGCTTCTGTGATTTACTGTGAGAGGTGACTGACTGAGTGTGCT	1644
Db	181	AGGGGAATTCGGCTTCTGTGACATCTCGGGGAGGTGACTTACACGTCGACCGGGTCTT	240
Qy	1645	GGATTAATAAATAGACTCTCTTCTCCGGAACTGGAAGGAGACATGTGCACTCATATAA	1704
Db	241	GGACAAAATAATGACTTCTTCTTCGGGAACCTTAAGGAGACCATGTGTAGCTCAAAAGAA	300
Qy	1705	CCCCATGATGGCCAGTCTTGTGACAAAGATGAGCTCAATGACAAAGAGCGGCCAG	1763
Db	301	TCCCATTTATGAGCAGTGTGTTTACCGGAGCGAGCTCACTGACAAAGAACCGCCAGAGAC	360
Qy	1764	GGTGGCACCCCACTTCAAGATGAGGCTCTGTGACGCTGTGAGATCTGAGGCTTAAGGA	1823
Db	361	GGTGGCACCCAGTTCAAGATGAGGCTCTGTGACGCTGTGAGATCTGAGGCTTAAGGA	420
Qy	1824	GGCTGCTATATCCGGTGTGATCAAGCCAAACGACGCCAAGCAGCCGGGTGCTTTGATGA	1883
Db	421	GGCCGCTAGCTGCGTGTGATCAAAACCCCAATGTGCAAAACGCCCCGCTGTTCAGCA	480
Qy	1884	GGTGTCTATCCGACATAGGTGAAGTAACTGGGACTGATGGAGAAATTCGGCGTCCGAG	1943
Db	481	GGTGTCTATCCGACACGAGTGAAGTAACTGGGAGCTGTTGGAAAACCTCGCGGTCCGAG	540
Qy	1944	AGCTGCTTTGCCATATGTCGTGCAAAATATAGGCTTTCCTGTGACAGAGTCAAGTCACTGTG	2003
Db	541	AGCGGGCTTTGGCTATGTCGCGCAATATACAGAGCTTTCCTGTGCAAAAGTACAACTCTGTG	600
Qy	2004	CCGACAGACATGGCCATGTGGGCGAGAGCGGCCACGATGGTGTGGCTGTGGTCAG	2063
Db	601	CCGACAGACATGGCCACGTCGGGCGAGAGCGGCGCGAGATGGGTGGCTGTGGTCG	660
Qy	2064	ACACTTGGCTACCAAGCCCAAGATACAAATATGGGCGAGAGTAAAGATCTTCATCCGATT	2123
Db	661	ACACTTGGCTACCAAGCCCAAGATACCAAGATGGGCGAGAGCAAGATCTTCATCCGATT	720
Qy	2124	CCCCAAGACCTTATTTTCCACAGAGAGATCTCCGTGAAGTCCGGGCGCAGAGTCTAGCCAC	2183
Db	721	TCCCAAGACCTTGTTCCTGACAGAGATGCTCCCTGGAAGTCCGGCGGACAGACTGTGGCAC	780
Qy	2184	CAGATGCTCAGGCGGCTGGAAGGGCTTTGCA-TTGGCGACAGAAATTTCTCCGGGTGAAGC	2242

Db 781 AAGATCCAGGCTGTCTGAGGGGCTTTCACCTGGCGCGCAAGATTCTCCCGCTTAAGA 840  
 QY 2243 GATCAGCCATCTGTATCCACT-CATGGTGGCTGGCAGACTGGG--CCGAGGAGAGCAG 2299  
 Db 841 GATCAGCCATCTGTATCCACTGGCGGGGGCTGGAACACTGGGGCCGAGAGAGCAG 900  
 QY 2300 CCAAGAGGAGTGGGCGAGCCCAACCATCCGTC 2332  
 Db 901 CCCACAGCAATGGGGGGCGGCAAGACCCCTTC 933

## RESULT 15

BQ958694

1068 bp mRNA

linear

EST 21-AUG-2002

LOCUS AGENCOURT\_10031162 NIH\_MGC\_40 Homo sapiens CDNA clone IMAGE:6479885  
 DEFINITION 5', mRNA sequence.

ACCESSION BQ958694  
 VERSION BQ958694.1 GI:22374172  
 KEYWORDS EST.

## SOURCE

human.

## ORGANISM

Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 AUTHORS NIH-MGC http://mgi.mgi.nhl.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@bbs-remail.nih.gov  
 Tissue Procurement: DCTD/DPF  
 CDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNLN at:  
 http://image.llnl.gov

Plate: LNCM2658 row: m column: 06  
 High quality sequence stop: 635.  
 Location/Qualifiers  
 1..1068

## FEATURES

source

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6479885"  
 /clone\_lib="NIH\_MGC\_40"  
 /tissue\_type="carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: prostate; Vector: pOT87; Site\_1: XhoI;  
 Site\_2: EcoRI; CDNA made by oligo-dr priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCAGGAG(G). Library constructed by  
 Ling Hong in the laboratory of Gerald M. Rubin (University  
 of California, Berkeley) using ZAP-CDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."  
 BASE COUNT 229 a 325 c 313 g 200 t 1 others  
 ORIGIN

Query Match 18.6%; Score 679.4; DB 14; Length 1068;  
 Best Local Similarity 85.0%; Pred. No. 3.5e-163;  
 Matches 806; Conservative 0; Mismatches 137; Indels 5; Gaps 4;

QY 2108 AGATCTTCATCCGATTCCTCCCAAGACTTATTTGGCAGAGAGACTCCCTGGAGTCCGGC 2167  
 Db 1 AGATCTTCATCCGATTCCTCCCAAGACTTATTTGGCAGAGAGACTCCCTGGAGTCCGGC 60  
 QY 2168 GGCAGACTTAGCCACCAAGATCCAGGGCGCTGGAGGGCTTTTCATTTGGCAGAGAAAT 2227  
 Db 61 GGCAGAGCTTGGCCACAAAGATCCAGCTGCGAGAGGGCTTTCACTGGCGGCAAGAAAT 120  
 QY 2228 TTCTCCGGGGAAGCGATGAGCATCTGTATCCAGTATGTCGCGTGGCAGACTGGGCC 2287  
 Db 121 TTCTCCGGGGAAGCGATGAGCATCTGTATCCAGTATGTCGCGTGGCAGAACTGGGCC 180

QY 2288 GGAGAGAGCAGCAGCAAGAGAGTGGGCGAGCCAGACCATCCGTCGACATCCGTGGCT 2347  
 Db 181 GGAGAGAGCAGCAGCAAGAGAGTGGGCGAGCCAGACCATCCGTCGACATCCGTGGCT 240  
 QY 2348 TCATTTTGGCCATTCACCCCGTGGCTGAGAAATGCTTTCTTTGGACACGACGCGC 2407  
 Db 241 TCATCTTGGCCAGCCCGCCCTGCGCGAGAACGCTTTCTCTGGACCATGTGGCA 300  
 QY 2408 CCTCATTTTGGCTTAACCTGAGAGGCGGCAACTGCCCCGGAATGTTTGGACACTCTGGC 2467  
 Db 301 CCTCATTTTGGCTTAACCTGAGAGGCGGCAACTGCCCCGGAATGTTTGGACACTCTGGC 360  
 QY 2468 CCACAGCCCGCAGCTGCTGAGAGAGGCGCTGAGAACTGCTCGGGAAGTGGCAAGAA 2527  
 Db 361 CCACAGCCCGCAGCTGCTGAGAGAGGCGCTGAGAACTGCTCGGGAAGTGGCAAGAA 420  
 QY 2528 ACATGCTGTGAAGTACTGCGGAGCATCAGCCCTGAGTGGAGACAGCAGCTGACGAAA 2587  
 Db 421 ACATGCTGTGAAGTACTGCGGAGCATCAGCCCTGAGTGGAGACAGCAGCTGACGAAA 480  
 QY 2588 AGCGGTGGCTAGTGAATTTTCAAGGCGCAAGAGACACTACCCCGAGAGTCCCA 2647  
 Db 481 AGCGGTGGCTAGTGAATTTTCAAGGCGCAAGAGACACTACCCCGAGAGTCCCA 540  
 QY 2648 GACTCTTCATTAAGACAGAGGCTTGGACAGAGAGATCAGCCCGAGAGTGTTCATCT 2707  
 Db 541 GACTCTTCATTAAGACAGAGGCTTGGACAGAGAGATCAGCCCGAGAGTGTTCATCT 600  
 QY 2708 TGGGCTCTGAACCCATCCAGTATGCGCGTGGTGAATAATACGACCGTAAAGGTTACA 2767  
 Db 601 TGGGCTCTGAACCCATCCAGTATGCGCGTGGTGAATAATACGACCGTAAAGGTTACA 660  
 QY 2768 AGCCTGCGCCCGGAGAGTGTGCTCAAGGCGCCAGTGTGTCATGTGGAGATGCTA 2827  
 Db 661 AGCCTGCGCCCGGAGAGTGTGCTCAAGGCGCCAGTGTGTCATGTGGAGATGCTA 720  
 QY 2828 AAGTCAAGCAGAGATTTGATTATGCAACCTTAACCGGAATCTGTGACAGTGGAGTGG 2887  
 Db 721 AAGTCAAGCAGAGATTTGATTATGCAACCTTAACCGGAATCTGTGACAGTGGAGTGG 780  
 QY 2888 ATAG-CCATTTTGGCTTACAGTGCAGCGTGAAG--ACAACAGCAGAGAGATGTTGG 2944  
 Db 781 ATAG-CCATTTTGGCTTACAGTGCAGCGTGAAG--ACAACAGCAGAGAGATGTTGG 840  
 QY 2945 TGCTGAGAGATCA-TGTGATCGAGACACTAACAAGAGGCGCTCACTGCTGACC-G 3002  
 Db 841 TGCTGAGAGATCA-TGTGATCGAGACACTAACAAGAGGCGCTCACTGCTGACC-G 900  
 QY 3003 CGTGAACAATATCAACATCAACAGGCGAGCATACGTTTGGAGGGGG 3050  
 Db 901 CGTGAACAATATCAACATCAACAGGCGAGCATACGTTTGGAGGGGG 948

Search completed: December 9, 2002, 15:10:49  
 Job time: 3255 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2002, 15:10:56 : Search time 67 Seconds

(without alignments)  
2076.323 Million cell updates/sec

Title: US-09-893-371-1

Perfect score: 5398  
Sequence: 1 MRYRASALGSDGVAVTMESEA.....LITKAKNGHLAVVAPRLNSR 1044

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5216	96.6	1063	22	AAAM39991
2	5165.5	95.7	1050	22	ABG10171
3	3369	62.4	935	22	ABG10172
4	2378	44.1	537	22	AAAM1777
5	1800.5	33.4	1011	22	ABM71113
6	1782	33.0	700	22	AAAB6415
7	1770.5	32.8	697	22	AAAB6416
8	1738.5	32.2	1026	22	AAU23125
9	1714.5	31.8	1098	23	AAU97344
10	1685	31.2	909	22	AAU23128

11	1565	29.0	1089	23	ABB97258
12	1412.5	26.2	759	22	ABG20610
13	1362	25.2	2167	22	ABB60369
14	1330.5	24.6	512	21	ABAB4261
15	1328	24.6	2129	22	ABB62828
16	1301.5	24.1	1493	21	AAAG4839
17	1301.5	24.1	1544	21	AAAG4838
18	1294	24.0	1483	21	AAAG4864
19	1284.5	23.8	2058	23	ABB97219
20	1272.5	23.6	2424	22	ABB58924
21	1269	23.5	1505	21	AAAG3043
22	1269	23.5	1556	21	AAAG3048
23	1261.5	23.4	1495	21	AAAG3044
24	1241.5	23.0	2048	22	AAE11891
25	1241.5	23.0	2057	22	AAE11890
26	1223.5	22.7	1754	21	AAAG5240
27	1223.5	22.7	1769	21	AAAG5240
28	1223.5	22.7	1769	21	AAAG5240
29	1218	22.6	1839	21	AAAG5240
30	1216	22.5	764	22	AAAM80123
31	1208	22.4	697	17	AAAM00603
32	1191	22.1	753	22	AAAM79139
33	1191	22.1	1939	23	ABB77096
34	1187.5	22.0	1453	22	AAAM39213
35	1186	22.0	569	21	AAAB57005
36	1183.5	21.9	1469	22	AAAM39214
37	1182	21.9	1120	21	AAAY94292
38	1181.5	21.9	631	17	AAAM00604
39	1181	21.9	1879	22	AAAM25750
40	1174.5	21.8	1960	22	AAAM78854
41	1174.5	21.8	1963	22	AAAM78854
42	1172.5	21.7	2056	22	ABB59344
43	1171	21.7	1988	22	AAAM40999
44	1171	21.7	1988	22	AAAM41000
45	1169.5	21.7	1972	17	AAAM00024

#### ALIGNMENTS

AAAM39991	RESULT 1
ID	AAAM39991 standard; Protein: 1063 AA.
XX	AAAM39991;
AC	22-OCT-2001 (first entry)
XX	Human polypeptide SEQ ID NO 3136.
DT	Human polypeptide SEQ ID NO 3136.
DE	Human: noctropic; immunosuppressant; cytosolic; gene therapy; cancer;
XX	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW	chemokine; thrombolytic; drug screening; arthritis; inflammation;
XX	leukaemia.
OS	Homo sapiens.
XX	WO200153312-A1.
PN	26-JUL-2001.
XX	26-DEC-2000; 2000WO-US34263.
PF	21-JAN-2000; 2000US-0488725.
XX	25-APR-2000; 2000US-0552317.
PR	09-JUL-2000; 2000US-0598042.
PR	19-JUL-2000; 2000US-0620312.
PR	03-AUG-2000; 2000US-0653450.
PR	14-SEP-2000; 2000US-0662191.
PR	19-OCT-2000; 2000US-0693036.
PR	29-NOV-2000; 2000US-0727344.

Novel human protei  
Novel human diagno  
Drosophila melanog  
Human ORF2415  
Drosophila melanog  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Arabidopsis thalia  
Drosophila protei  
Drosophila melanog  
Arabidopsis thalia  
Arabidopsis thalia  
Angiogenesis assoc  
Angiogenesis assoc  
Arabidopsis thalia  
Arabidopsis thalia  
Caenorhabditis ele  
Human protein SEQ  
Human non-conventi  
Human protein SEQ  
Human alpha-myosin  
Human polypeptide  
Human prostate can  
Human polypeptide  
Helianthus annuus  
Murine myosin VII  
Human protein sequ  
Human protein SEQ  
Drosophila melanog  
Human polypeptide  
Human polypeptide  
Smooth muscle myos



XX (HYSE-) HYSEQ INC.  
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 XX Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QH, Zhou F, Goodrich R, Drmanac RT;  
 XX WPI: 2001-442253/47.  
 DR N-PSDB: AAI59147.  
 XX Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 XX  
 PS Example 4: SEQ ID NO 3136: 10078pp; English.  
 XX The invention relates to human nucleic acids (AA157798-AA161369) and  
 CC the encoded polypeptides (AAM38642-AA442213) with nucleotide,  
 CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC centralised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 CC  
 XX Sequence 1063 AA:  
 SQ  
 Query Match 96.6%; Score 5216; DB 22; Length 1063;  
 Best Local Similarity 96.6%; Pred. No. 0;  
 Matches 1003; Conservative 22; Mismatches 13; Indels 0; Gaps 0;

QY 547 LFRNLKTMSSMNPIMAOFCDSKSELDKKRPETVATQFMASLLOLVEITRSKPEAYINC 606  
 DB 566 LFRNLKTMSSMNPIMAOFCDSKSELDKKRPETVATQFMASLLOLVEITRSKPEAYINC 625  
 QY 607 IKPNDAKOPRPFDEVLIRHOKYTLGLMENLRYRAGFAVRKKEALQORXSLCPETWM 666  
 DB 626 IKPNDAKOPRPFDEVLIRHOKYTLGLMENLRYRAGFAVRKKEALQORXSLCPETWM 685  
 QY 667 WAGRPDDGVAIVLRHLYGKPEEYKMGRTKLEIRPEKTLFATEDSLRQSLATKIQAM 726  
 DB 686 WAGRPDDGVAIVLRHLYGKPEEYKMGRTKLEIRPEKTLFATEDSLRQSLATKIQAM 745  
 QY 727 RGFHMRKQFLRYKRSALCTOSMWRGTLGRKAKAKRMAQTIRRLRGFTLRAPRCPEN 786  
 DB 746 RGFHMRKQFLRYKRSALCTOSMWRGTLGRKAKAKRMAQTIRRLRGFTLRAPRCPEN 805  
 QY 787 AFELDHVRASFLNLRQLPRNVLDTSWPTPPALREASELLRELCKMWMKRCSTSP 846  
 DB 806 AFELDHVRASFLNLRQLPRNVLDTSWPTPPALREASELLRELCKMWMKRCSTSP 865  
 QY 847 EMKQLOQKAVASEIFKGRKDNYPQSVPRLEISTRLGTEISPRVLOSISPIQYAPV 906  
 DB 866 EMKQLOQKAVASEIFKGRKDNYPQSVPRLEISTRLGTEISPRVLOSISPIQYAPV 925  
 QY 907 VKYDRKGYRPRQOLLTFSAYIVYDAKVKORIDYANLTGISVSLDSFLVHVQRED 966  
 DB 926 VKYDRKGYRPRQOLLTFSAYIVYDAKVKORIDYANLTGISVSLDSFLVHVQRED 985  
 QY 967 NKQKGVNLOSDBVIELTKTALSADRVNNININOSITFAGGPGDGIIDTSGSELLI 1026  
 DB 986 NKQKGVNLOSDBVIELTKTALSADRVNNININOSITFAGGPGDGIIDTSGSELLI 1045  
 QY 1027 TKAKNGHLAVAPRLNSR 1044  
 DB 1046 TKAKNGHLAVAPRLNSR 1063

RESULT 2  
 ABG10171  
 ID ABG10171 standard; Protein: 1050 AA.  
 XX  
 AC ABG10171;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #10162.  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PT Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI: 2001-639362/73.  
 DR N-PSDB: AAS74358.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -

XX Claim 20; SEQ ID NO 40530; 103pp; English.  
PS The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes.  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG3037 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1050 AA:  
Query Match 95.7%; Score 5165.5; DB 22; Length 1050;  
Best Local Similarity 95.8%; Pred. No. 0;  
Matches 999; Conservative 24; Mismatches 19; Indels 1; Gaps 1;  
QY 7 ALGSDGVRVTMESALTRADRVGVDFVLENTSEAFIENLRFRRENLIYTYIGPVLY 66  
DB 12 ALGSDGVRVTMDSLTRADRVGVDFVLENTSEAFIENLRFRRENLIYTYIGPVLY 71  
QY 67 SVNPYRLQIYSRQHMERYRGVFEVPPHLPAAVDTYVYALRTERDQAVMISGSGAG 126  
DB 72 SVNPYRLQIYSRQHMERYRGVFEVPPHLPAAVDTYVYALRTERDQAVMISGSGAG 131  
QY 127 KTEATKRLLOFYAETCPAPRGGAVERDLOSNPVLEAFNAKTLRNNSRFGKYMYVO 186  
DB 132 KTEATKRLLOFYAETCPAPRGGAVERDLOSNPVLEAFNAKTLRNNSRFGKYMYVO 191  
QY 187 PDFGAPVGVGHILSYLLEKSRVHQNNGERNFHFYQLLEGGEEETLRLLEENPOSYL 246  
DB 192 PDFGAPVGVGHILSYLLEKSRVHQNNGERNFHFYQLLEGGEEETLRLLEENPOSYL 251  
QY 247 YLVAGQCAKVSINDKSMKVMKALSYIDFTEDEVEDLSTVASVHLGNIHFAEDDS 306  
DB 252 YLVAGQCAKVSINDKSMKVMKALSYIDFTEDEVEDLSTVASVHLGNIHFAEDDS 311  
QY 307 NAOYTTEMLKYLTRLGVECTTLREALTHRKTIKAGEELISPLNLEQAAVARDALAKAV 366  
DB 312 NAOYTTEMLKYLTRLGVECTTLREALTHRKTIKAGEELISPLNLEQAAVARDALAKAV 371  
QY 367 YSRFTFTMLVRKINSLSKDAESPWSRSTVVLGLDITGFVFOHNSFEQFCINCYCNEL 426  
DB 372 YSRFTFTMLVRKINSLSKDAESPWSRSTVVLGLDITGFVFOHNSFEQFCINCYCNEL 431  
QY 427 QQLFIELTLKSEGEYEAGIAMEPVGYFNKKIICDLVEEKFKGIISLDEECRLPGEAT 486  
DB 432 QQLFIELTLKSEGEYEAGIAMEPVGYFNKKIICDLVEEKFKGIISLDEECRLPGEAT 491  
QY 487 DLTFLEKLEDFVVKHPHPLTHKLDQKTRKSLDGEFFLLIYAGGVNYSVNGFLDKNDL 546  
DB 492 DLTFLEKLEDFVVKHPHPLTHKLDQKTRKSLDGEFFLLIYAGGVNYSVNGFLDKNDL 551  
QY 547 LFRNLKETMCSMPIMACFSELSDKKRPETVATQFKKSLQLVILSKPEAVYRC 606  
DB 552 LFRNLKETMCSMPIMACFSELSDKKRPETVATQFKKSLQLVILSKPEAVYRC 611  
QY 607 IKPDNAKOPGFDEVLLRHQVYGLMENLVRRAQFAVRRKYEAFLQRYKSLCPETWPM 666  
DB 612 IKPDNAKOPGFDEVLLRHQVYGLMENLVRRAQFAVRRKYEAFLQRYKSLCPETWPM 671

DB 612 IKPDNAKOPGFDEVLLRHQVYGLMENLVRRAQFAVRRKYEAFLQRYKSLCPETWPM 671  
QY 667 WAGRPQDVAALVRLHGLGKPEEYKMGRTKIFIRPKTLFATEDSLVBRQSLATKIQAM 726  
DB 672 WAGRPQDVAALVRLHGLGKPEEYKMGRTKIFIRPKTLFATEDSLVBRQSLATKIQAM 731  
QY 727 RGFHMRQKFLVKKRSALICIQSMWRGTLGRRAKRAKMAQOTIRRLINGFTLRHSPREN 786  
DB 732 RGFHMRQKFLVKKRSALICIQSMWRGTLGRRAKRAKMAQOTIRRLINGFTLRHSPREN 791  
QY 787 AFPLDHAASFLNLRDLPKNVLDTSWPTPPALREASSELLRELCKMNMWVKYRSISP 846  
DB 792 AFPLDHAASFLNLRDLPKNVLDTSWPTPPALREASSELLRELCKMNMWVKYRSISP 851  
QY 847 EMKQLOOKAAVASELFFKKKNYPOSVPRFLFSTRGLTEESIPRLOSIGSEPIQYAVPV 906  
DB 852 EMKQLOOKAAVASELFFKKKNYPOSVPRFLFSTRGLTEESIPRLOSIGSEPIQYAVPV 911  
QY 907 VKYDRKGYKPPRRQLLTPS-AVIVEDAKVKQRIIDYANLTGISVSSLSLFLVHORE 965  
DB 912 VKYDRKGYKPPRRQLLTPS-AVIVEDAKVKQRIIDYANLTGISVSSLSLFLVHORE 971  
QY 966 DNKQKGDVVLQSDHYIEFTLTALSDADRVNNTINQGSITFAGRGRCIIDFTSGSELL 1025  
DB 972 DNKQKGDVVLQSDHYIEFTLTALSDADRVNNTINQGSITFAGRGRCIIDFTSGSELL 1031  
QY 1026 IYKAKNGHLAVVAPRLNRSR 1044  
DB 1032 IYKAKNGHLAVVAPRLNRSR 1050  
RESULT 3  
ABG10172  
ID ABG10172 standard; Protein: 935 AA.  
XX  
AC ABG10172;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #10163.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PE 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
XX  
NR N-PSDB: NAA574359.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity -  
PS Claim 20; SEQ ID NO 40531; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 935 AA:

Query Match 62.4%; Score 3369; DB 22; Length 935;  
 Best Local Similarity 86.0%; Pred. No. 6.2e-278;

Matches 662; Conservative 22; Mismatches 20; Indels 66; Gaps 6;

QY 190 KGAPVGHILSYLLEKSRVHONHGERNFVYQLLEGGEETLRRLGERNPQSYLYIV 249  
 DB 82 KGAPVGHILSYLLEKSRVHONHGERNFHIFQYLLGGEETLRRLGERNPQSYLYIV 141  
 QY 250 KGCCAVKSSINDSKDMKVMKALSVIDFTPEDEVEDLSTVASYLHGNHFADEDSMAQ 309  
 DB 142 KGCCAVKSSINDSKDMKVMKALSVIDFTPEDEVEDLSTVASYLHGNHFADEDSMAQ 201  
 QY 310 VTTENOLKYLTRLGVEGTTLRREALTHRKIIAKGELLSPINLEQAAYARDALAKVYSR 369  
 DB 202 VTTENQVKYTLRLSYEGSTLRALHRIKIIAGCELLSPINLEQAAYARDALAKVYSR 261  
 QY 370 TTFWLVKIRKIRSLASK---DAEESPR---STTVLGLDLYGFEVPHNSFEQFCINYC 423  
 DB 262 TTFWLVGKIRKIRSLASKRPARPPLSWSCVPTFPLMLLPY-----SFEQFCINYC 313  
 QY 424 EKLOQLFIETLKSDEEYEAEGIAPEYQFNKIIICOLVERKFKGIIISIDECLARG 483  
 DB 314 EKLOQLFIETLKSDEEYEAEGIAPEYQFNKIIICOLVERKFKGIIISIDECLARG 373  
 QY 484 EATDLTFLEKLEDTYKRNHFLTHKLADOKTRKSLDRGEFRLHAGEVTVSGFLDKN 543  
 DB 374 EATDLTFLEKLEDTYKRNHFLTHKLADOKTRKSLDRGEFRLHAGEVTVSGFLDKN 433  
 QY 544 NDLLFNKLKETWCSSNRPIMACQFDSKSKKRPETVATQKMSLLQLVELIRSKEPAY 603  
 DB 434 NDLLFNKLKETWCSSNRPIMACQFDSKSKKRPETVATQKMSLLQLVELIRSKEPAY 493  
 QY 604 IFCIRPNDAKQ-----FGRDEVLIIRHGVYUUGLGMENLRYRBRGPFAYR 646  
 DB 494 VFCIRPNDAKQCGDSKQRPADHGAIRPTGRDEVLIIRHGVYUUGLGMENLRYRBRGPFAYR 541  
 QY 647 RRYEAFLOQKSKLCEPTPMAGRGODGVAVLVRLHLYKRPBEYKMGRTKIFIRPKTLFA 706  
 DB 542 RRYEAFLOQKSKLCEPTPMAGRGODGVAVLVRLHLYKRPBEYKMGRTKIFIRPKTLFA 601  
 QY 707 TEDSELEVRSLATKIQAAHNGFIMHROKFLARKRSALICIQSMWRSTLCRRKAARKKMAAQ 766  
 DB 602 TEDSELEVRKSL-----ALICIQSMWRSTLCRRKAARKKMAAQ 638  
 QY 767 TIRRLIRGFIIRHSPRCENAFELDHVASEFLNLRLRPLNVLDTSPPTPPALREASE 826  
 DB 639 TIRRLIRGFIIRHSPRCENAFELDHVASEFLNLRLRPLNVLDTSPPTPPALREASE 698  
 QY 827 LLRELCKNNMVKYGRSISPEKQOOLQAAVASLTFKGRKDNYPQSVPRLEISTRIGTEE 886  
 DB 699 LLRELCKNNMVKYGRSISPEKQOOLQAAVASLTFKGRKDNYPQSVPRLEISTRIGTEE 758  
 QY 887 ISPRVLQSLGSEPIQYAVPVVKYDRKGYKRRPQLLLTPSAVVIYEDAKV 936

DB 759 ISPRVLQSLGSEPIQYAVPVVKYDRKGYKRRPQLLLTPSAVVIYEDAKV 808

RESULT 4

ID AAM41777 standard; Protein: 537 AA.

XX AAM41777;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 6708.

XX Human; nocotropic; immunosuppressant; cytoskeletal; gene therapy; cancer;

KW peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

XX Leukaemia.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0553317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HISE-) HISEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.

XX N-PSDB; AAI60933.

XX Novel nucleic acids and polypeptides, useful for treating disorders

XX such as central nervous system injuries -

XX Example 2: SEQ ID NO 6708; 10078bp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and

XX the encoded polypeptides (AAM38642-AAM42213) with nocotropic,

XX immunosuppressant and cytoskeletal activity. The polynucleotides are useful

XX in gene therapy. A composition containing a polypeptide or polynucleotide

XX system, such as peripheral nervous injuries, peripheral neuropathy and

XX localized neuropathies and central nervous system diseases, such as

XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

XX utilisation of the activities such as: immune system suppression,

XX CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

XX assays for receptor activity, cancer diagnosis and therapy, drug screening,

XX CC C.N.S disorders.

XX Note: The sequence data for this patent did not form part of the printed

XX specification.

XX Sequence 537 AA:

Query Match 44.1%; Score 2378; DB 22; Length 537;  
 Best Local Similarity 96.7%; Pred. No. 9.1e-194;

Matches	462:	Conservative	8:	Mismatches	8:	Indels	0:	Gaps	0:						
Qy	7	ALGSDGVRTMESALTARDRVGVDFVLLNFNTSEAFIENLRFRRENLYTYIGPVLV	66												
Db	60	ALGSDGVRTMESALTARDRVGVDFVLLNFNTSEAFIENLRFRRENLYTYIGPVLV	119												
Qy	67	SVNPRDLOIYSRQHMERYRGVSTFEVPHLFVADYTYRALKTERRQOAVMISGSGAG	126												
Db	120	SVNPRDLOIYSRQHMERYRGVSTFEVPHLFVADYTYRALKTERRQOAVMISGSGAG	179												
Qy	127	KTEATKRLLOFYAETCPAPERGAVERDRLQSNPYLEAFGNKTLRNDSSRFKGYMVO	186												
Db	180	KTDATKRLLOFYAETCPAPERGAVERDRLQSNPYLEAFGNKTLRNDSSRFKGYMVO	239												
Qy	187	FDGKAPVGHILSYLLEKSRVHONHGERNFHFVYQLLEGSEETLRRLGLERNPOSYL	246												
Db	240	FDGKAPVGHILSYLLEKSRVHONHGERNFHFVYQLLEGSEETLRRLGLERNPOSYL	299												
Qy	247	YLVKGOCACKVSSINDSKSMWKAKALSYIDFTEDEVEDLLSTVASVHLGNHFAADEDS	306												
Db	300	YLVKGOCACKVSSINDSKSMWKAKALSYIDFTEDEVEDLLSTIASVHLGNHFAANES	359												
Qy	307	NAOYVTEMLQKTLRLGLGEGTTLREALTHRKIIAKGSELLSPNLBQAAYARADALAKAV	366												
Db	360	NAOYVTEMLQKTLRLGLGEGTTLREALTHRKIIAKGSELLSPNLBQAAYARADALAKAV	419												
Qy	367	YSRFTFWLVRKINRSLSKDAESPMSRSTVYGLDIYGFVFOHNSFEQFCINCNEL	426												
Db	420	YSRFTFWLVRKINRSLSKDAESPMSRSTVYGLDIYGFVFOHNSFEQFCINCNEL	479												
Qy	427	QOLFIETLTKSBOEYEAEGTAMEPVOYFNKKIICDLVEKFKGIISLDEECLRPE	484												
Db	480	QOLFIETLTKSBOEYEAEGTAMEPVOYFNKKIICDLVEKFKGIISLDEECLRPE	537												
RESULT 5															
AB71113	ID	AB71113	standard; Protein: 1011 AA.												
XX	AC	AB71113;													
XX	DT	26-MAR-2002 (first entry)													
XX	DE	Drosophila melanogaster polypeptide SEQ ID NO 40131.													
XX	KM	Drosophila: developmental biology; cell signalling; insecticide;													
XX	OS	Drosophila melanogaster.													
XX	PN	WO200171042-A2.													
XX	PD	27-SEP-2001.													
XX	PE	23-MAR-2001; 2001WO-US09231.													
XX	PR	23-MAR-2000; 2000US-191637P.													
XX	PR	11-JUL-2000; 2000US-0614150.													
XX	PA	(PEKE ) PE CORP NY.													
XX	PI	Venter JC, Adams M, Li PWD, Myers EW;													
XX	DR	WPI: 2001-656860/75.													
XX	DR	N-PSDB: ABL15216.													
XX	PT	New isolated nucleic acid detection reagent for detecting 1000 or more													
XX	PT	genes from Drosophila and for elucidating cell signalling and cell-cell													
XX	PS	interactions -													
XX	CC	Disclousure: SEQ ID NO 40131; 21pp + Sequence Listing; English.													
XX	CC	The invention relates to an isolated nucleic acid detection reagent													

CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC	sequences (AB101840-AB116175) and the encoded proteins
CC	(AB57737-AB72072).
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	Sequence 1011 AA:
Qy	Query Match 33.4%; Score 1800.5; DB 22; Length 1011;
Qy	Best Local Similarity 39.6%; Pred. No. 4.5e-144;
Matches	426; Conservative 169; Mismatches 341; Indels 139; Gaps 29;

Qy	28	GVDFVLLNFNTSEAFIENLRFRRENLYTYIGPVLSVNPYDLOIYSRQHMERYRG	87						
Db	8	GVDFVLLDQVSM-KFMDNLKRRFONGSIYTYIGEVCSNMPYROMNIYGETIRKYG	66						
Qy	88	VSFYVPPHPLFVADTYVRALRTERRQOAVMISGSGAGKTEATKRLLOFYAETCPAP	145						
Db	67	RELFEPAHLFRAIDSAVRVLKOROOTCILISGSGAGKTEASKITMKYIAAVTNAGQ	126						
Qy	146	--ERGAVRDLQSNPYLEAFGNAKTLRNDSSRFKGYMVOVDFGAPVGHILSYL	202						
Db	127	NEIER--VKNVLIQSNMILETFGNAKTLRNDSSRFKGYMIEDYADAPVGIITNVL	183						
Qy	203	LEKSRVHONHGERNFHFVYQLLEGSEETLRRLGLERNPOSYLTVKGOCACKVSSINDK	262						
Db	184	LEKSRVHONHGERNFHFVYQLLEGSEETLRRLGLERNPOSYLTVKGOCACKVSSINDK	240						
Qy	263	SDMKVMKALSYIDFTEDEVEDLLSTIASVHLGNHFAADEDSNAOYVTEMLQYLFR	322						
Db	241	SDYKTCNAFKTLGFTSTDEVOTIMRTIAVHLGNVEFTIED-ELVISNKHLSSTAKL	299						
Qy	323	LGVEGTLREALTHRKIIAKGSELLSPNLBQAAYARADALAKAVYSRFTFWLVRKINRS	382						
Db	300	LOVTELTSTALTKRVIAAGVNMQKDHNAQAEYKDALAKAIVDRLEFTWIIISRINRAI	359						
Qy	363	ASKDAESPMSRSTVYGLDIYGFVFOHNSFEQFCINCNELQOLFIETLTKSEOEY	442						
Db	360	LFKRSKTYQA-RENSIVAGVLDYIGFEIIFDINSFEQFCINCNELQOLFIETLTKSEOEY	418						
Qy	443	BAEGTAMEPVOYFNKKIICDLVEKFKGIISLDEECLRPGCATDLTFLEKLEDVYKPP	502						
Db	419	OREGIEWTNIEYFNKKIICDLVEOPHKGITAIMDACLGVGAVTDOTLLGAMDKMLSKHP	478						
Qy	503	HELTHTLADQKTRKSL-DRGEFRLHYAGEVYVTGFLDKNNDLFR-----NLKET	554						
Db	479	HYTSROL--KPTDKELKHREDRITHYAGDVYINNGFIEKKKDTLYQDFKRLNLSKDA	536						
Qy	555	MCSMNPTMAOCSELSDDKRPETVATQFKMSLLQVLEILRSKPAVIRICIKRNDAKQ	614						
Db	537	NLSEMPPEGADODIKT---TKRPLTAGTLFORSMADEVTLTKKEPFEVRCIKRNDLKS	592						
Qy	615	PGRPDEVILRHQVXYLGLMENLRVRAGAVRKYKEAPLQKSKLCPETWPMW-AGRPD	673						
Db	593	STVFDEEREYHGVRYLGLLENLRVRAGVHQRKDFELRKMSQYTWPMFNRGSDSD	652						
Qy	674	GVAIVLRHLYKPEEYKMGKRTKIFTRPKTLFATDESLVRQSLATKIQAMRFGFWHQ	733						
Db	653	GVRLVLEKKF-AQDVKGGHTKIFIRSPRTLEALHQRNEMPHIVTLQKVRGMIVNR	711						
Qy	734	KFLVYKRSATICTOSWWRGTGLRRKAKRKAQAOTIRLRILRGILRHSPPCPENAEFFLDHV	793						
Db	712	NKKKK-----AAITIVRAKAKKLR-----	732						
Qy	794	RASFLNLRLROL--PRNVLD---TSWPTPPALREASELR---ELCKMNV-WKYRS	843						
Db	733	--SYOELANRLKAKOMRDYKSIOWPOPPLAGKRVKAKLHRMDFWRANMILHKYPRS	790						

XX

XX

PN WO200077197-A1.  
XX 21-DEC-2000.  
XX 01-JUN-2000; 2000WO-US14934.  
XX 11-JUN-1999; 99US-0138599.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA (ROSE) ROSEN C A.  
XX Rosen CA, Ruben SM, Komatsoulis GA:  
PI WPI: 2001-032312/04.  
XX Isolated nucleic acid molecule encoding a human secreted protein is  
PT used in preventing, treating or ameliorating a medical condition  
XX  
XX Disclosure: Page 519-521: 558pp; English.  
XX  
XX The invention relates to the isolation of genes AAF32757-F32803 encoding  
CC the human secreted proteins AAB64549-B64594. The sequence is used as a  
CC query sequence for doing BLASTX searches to identify homologous  
CC sequences. The genes and proteins are useful for preventing,  
CC ameliorating or treating medical conditions, e.g. by protein or gene  
CC therapy. The genes are isolated from a range of human tissues disclosed  
CC in the specification. The nucleic acids, proteins, antibodies and  
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:  
CC (a) cancer, e.g. breast and ovarian cancer, and other cancers of the  
CC adrenal gland, bone, bone marrow, breast, gastrointestinal  
CC tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's  
CC disease, allergies, autoimmune haemolytic anaemia, autoimmune  
CC thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis,  
CC rheumatoid arthritis and ulcerative colitis; (c) cardiovascular  
CC disorders such as myocardial ischaemia; (d) wound healing; (e)  
CC neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
CC infectious diseases such as viral, bacterial, fungal and parasitic  
CC infections.  
XX  
XX Sequence 697 AA:  
SQ

Query Match 32.8%; Score 1770.5; DB 22; Length 697;  
Best Local Similarity 49.9%; Pred. No. 9.2e-142;  
Matches 352; Conservative 118; Mismatches 190; Indels 45; Gaps 7;

QY 27 VGQVDFLENTSEAFTEFLRRPRENLITYIGPVLSVNPYRDQIYSRQHEHYR 86  
DB 15 IGVGMVLEPL-NEETFINNLKRRFDHSEITYIGSVISVNPRLSPITYSPEKEEYR 73  
QY 87 GVSFEVPHLEFAVDVTVYRALRTERRDQAVMISGESGAGTEATKRLQFYAETCPAPE 146  
DB 74 NNNFELSPHIALSDEAVRSLRDQDKCILITGESGAGTEASKLWMSYAAVCGKA 133  
QY 147 RCGAVRDLQSNPVYLEAFAGNAKTLANDNSSRRFGKMDYQDFPGKAPVCGHLLSTYLEKS 206  
DB 134 EVNOVKEDQLQSNPVYLEAFAGNAKTVANDNSSRRFGKMDIEFPFGKPLGQGVISNYLLEKS 193  
QY 207 RVVHONHGRNHFVFOYLEGGEETFLRLGLERNPOSFLYLVKGCQAVSSINPKSDMK 266  
DB 194 RYVKQPRGRNHFVFOYQLLSGASEELNKLKLEPRDSKRNLY-SLDSAKVNGVDANR 252  
QY 267 VNRKALSVIDFTEDEVELLSIVASVLHGINHFAADESNMQVTTENOLKYLTRLLGVE 326  
DB 253 TVRNAMQIVGFMDHEASVLAVALKGNIEFKPE-----SRNGGID 296  
QY 327 GTTLREALTHRKI-----IAKGEELSPLELDAQAAVARALAKAVYS 368  
DB 297 ESKIDKMSXKRFVNCRHXSSISRRAFSPRTVEPEKYSTLVNQAAYARALANLVS 356  
QY 369 RFTFVLVRKINSLASKDAESPMSWSTYVLGLDITGFVFOHNSPEOCINVCNKLOO 428  
DB 357 RLFSMLVNRINESIKAKQTRVKRK-----VMGVLDITGFIPEQNSFEQFIINVCNKLOO 411

QY 429 LFIETLKSEQEEYEAEGTAMEPVQYFNKKIICDLVEEKFKGIISLDECLRPGEATDL 488  
DB 412 IFIEELTKEQEEYIREDIEMTHIDYFNNAIICDLLENTNGLAMLDECLRPGETVDE 471  
QY 489 TFELEKLEDVFKPHPHFLTH--KLADQKTRKSLDRGEFLHLHYAGEVTVSGTGLDKNNDL 546  
DB 472 TFELEKLNQVCATHQHEESMSKCSRFPLNDTSLPHSCFRLQHTAGKVLQYEGVDNNDL 531  
QY 547 LFRNLKETWCSMNPIMACFPKSELS--DKRRPETVATQFKMSLQLEILRSKPEAYI 604  
DB 532 LYRDLQAMMKASHALIKSLFPEGNPAKINLKRPTFAGSQFKASVATLKNLOTKNPNYI 591  
QY 605 RCIKPNDAKQGRFDEVLIRHGVKYLGLMENTLVRPAGFAYRRKYEATLQRYKSLCPETW 664  
DB 592 RCIKPNDKKAHIFNEALVCHQIRYLGLLENVVRBRAGYAFRQAYPCLEERYKMLCKQTM 651  
QY 665 PMWAGRPQGVAVLVHNLGKPEEYKMGRTKFIIRPKLFATED 709  
DB 652 PHMKGPARGVEVLFNELEIPVEEYSFGRSKIFIRNPRLFLED 696

RESULT 8  
AAU23125  
ID AAU23125 standard; Protein; 1026 AA.  
XX  
XX AAU23125;  
AC  
XX  
DT 17-DEC-2001 (first entry)  
XX  
DE Novel human enzyme polypeptide #211.  
XX  
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;  
XX ligase; hyperproliferative disorder; immunodeficiency disorder;  
XX autoimmune disorder; neurological disorder; metabolic disorder;  
XX inflammatory disorder; cardiovascular disorder; reproductive disorder;  
XX blood related disorder; infectious disorder; cyostatic; anti arthritic;  
XX nephrotropic; anticoagulant.  
XX  
XX Homo sapiens.  
XX  
XX WO200155301-A2.  
PN  
PD 02-AUG-2001.  
XX  
XX  
PF 17-JAN-2001; 2001WO-US01239.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205115.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0232081.  
PR 14-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233053.  
PR 14-SEP-2000; 2000US-0233064.  
PR 21-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234977.  
PR 25-SEP-2000; 2000US-0234988.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0235837.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239335.  
PR 13-OCT-2000; 2000US-0239337.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
PR (HUMA-) HUMAN GENOME SCI INC.  
PR PA  
PR XX  
PR XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI: 2001-465566/50.  
XX  
DR N-PSDB; AAS40995.  
DR  
XX  
PT Novel polypeptides and polynucleotides useful for diagnosing,  
PT preventing, treating neural, immune system, muscular, reproductive,  
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous  
PT diseases  
XX  
XX  
PS Claim 11; SEQ ID NO 1121; 1180pp; English.  
XX  
XX The present invention relates to the isolation of novel human enzyme  
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences  
CC encoding them. The enzyme polypeptides of the invention may comprise the  
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,  
CC isomerases or ligases. The sequences of the invention are useful in the  
CC diagnosis, treatment, prevention and/or prognosis of a wide range of  
CC disorders including hyperproliferative disorders (e.g. cancer),  
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders  
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),  
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders  
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),  
CC blood-related disorders (e.g. haemophilia), reproductive disorders  
CC (e.g. infertility) and infectious disorders (e.g. Influenza). The  
CC polynucleotides of the invention can also be used in gene therapy.  
CC AAU2915-AAU23814 represent the novel human enzyme polypeptides of the  
CC invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 1026 AA:

```

Query Match      32.2%; Score 1738.5; DB 22; Length 1026;
Best Local Similarity 41.6%; Pred. No. 9e-139;
Matches 404; Conservative 155; Mismatches 339; Indels 73; Gaps 22;

QY 28 GVQDVLLENFTSEAFLENLRRLRRRENTIYTYIGPVLYSVNVPYRDLOIYSQIHMERRG 87
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 18 GKPDPVLLDQVYME-DFMKNLQLRFEKRIYTYIGEVLYSVNVPYDELLPYGEALTAARYG 76
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 88 VSFYEVPHLEAVADTVVRAALRTERRDOAVMISGSGAGKTATKRLLOFYETCPAPER 147
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 77 RELYRPHPLAVANAAAKKAMKHSRDCTIYSGSGAGKTASHIMQYIAAVINPSCR 136
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 148 GCA--VDRLLQSNVLEAFGNAKTLRNDSSRFEGKYMDVDFEGAPYGVGHISLYLLEK 205
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 137 AEVEYKFDVLLKSTCYLBAFGNARTNRNHNSSRFQKMDINDFKDPGIGHISLYLLEK 196
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 206 SRVYNHNGERFHFYQYLLGEGEETLRRLGLEENPOSYLLVAGQCAKV---SSI-ND 261
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 197 SKVLKHNGERFHFYQYLLRGSEDKOLHELRLERPAVYNFTHQGAGLIMTVSHALSD 256
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 262 KSDMKVMRKALSVIDFTEDEVEDLSIVAVLHGNINHPAEDSDNAQ-----VTTENQL 316
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 267 EQSHQAVNTEAMVIEFSPBEVSVHRIILAILHGNIEFEVEEGGLOKEGLAVAEALV 316
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 317 KYLRLLEGGTTLREALTHRKIIAKGEBLSP-LNLEQAAVARDALAKAVYSRTFTWLV 375
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 317 DHVAELTATPRDLVRLSLRTAFVAGSRELIKGHTAAEASVARDACAKAVQRLFEWV 376
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 376 RKINSLSKQDESPSMSTYVLGLDITGFEVFOHNSFEQFCINVCNKLQQLFIELTL 435
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 377 NMINSVMERGRDPRDGKDTYIGVLDITGFEVFNSEDFQFCINVCNKLQQLFIELTL 436
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 436 KSEOEVEAEGLTAMPEVVOYFNKKIICDLVEEKFGLIISLDECLRPGEATDLTFLEKLE 495
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 437 KQOEVEEVEEGTINQSVFENFATYDVLVERHKGILAVLDACSSAGITTRIFLQITLD 496
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 496 DTVKRPHFELHKLADQKTRKSLDRG-EFRLLYAGEVTVSYTGELDKNNLLFRNLKET 554
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 497 THNRHNLHTYSQIC--PTDKTMEFGDRFRKHYAGADVTVSEGVFIDKKRDLFDQFKRL 554
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 555 KSSSNMPTMAOCF-----DKSELSDKKREYVATOFKNSLLOLVLELSKEPAVYRRCIP 609
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 555 LYNSTDPLIRAMPDGOODITEVT--KRLTATGTLTKNSMVALVLEMSKEPFYVRCIAP 612
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 610 NDAKOPGREDFLETHNOVYUIGLIMENLRVRAGFAVRRKYEAFLOVRYKSLCEPTMP-MMA 668
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 613 NEDKTAAGKIDENHCHNOVAUYLLELVNVRKRGFASRQYSRFLRYKMTCEYTPNHL 672
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 669 GRPQGVAVLVHILGYKPEEYMGRTKIFIRPKTULFATEDSLEVRKOSLATKIQAAWRG 728
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 673 GSDKAASVALLERQHLQ-CDVAFGHSKILFRSPRTLVTLEQS----- 713
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 729 FIMWRKFLVRKSAICISQWMRGSTLGRKKAARKMAAOITRRLINGFILRHSPCEENAF 788
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 714 ---RARLPI--TVLLLOAMRGTLARMR-CRRRLRIYIMWFR---RHKVRA----- 758
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 789 FLUDHVASFLMLRLRPLRVNLDTSWPPPALREASELRELCKMKNMWKTCRSISPEM 848
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 759 HLAELQRRF--QAARQPLXGRDLWPLRPRAVLQRPQDCHALFCMRKARQLVKNIPSP 816
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 849 KOOLQOKAVASIEFKGKDNV--POSVPLRISTRLGTEEISPRVLOSL-----GSEP 899
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 817 MPQIKAKVAMAGALQRLQDWCRRAMARDYLSATDNPJTASSLFAQRLKTLRDKDGFCA 876
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 900 IOYAVPVYKDKKYPKRRROLLTPRSANVIVE---DAKVQRIDYANLTGISVSSLS 956
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 877 VLESSEHVRKVN-RFKIRRALLLTDQHLKLDPPQRYVMRAVPLEAVTGLSVTSGDQ 935
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 957 LFLVLHOREDN 967
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 936 LVVLAHARGDD 946

```

```

RESULT 9
AAU97544
ID AAU97544 standard; Protein; 1098 AA.
XX
AC AAU97544;
XX
DT 13-AUG-2002 (first entry)
XX
DE Human Myosin-II protein MYOLF.
XX
KM Human; phosphoinositide-binding protein; phosphatidyl acid; MYOLF;
XX phosphatidic acid; signal transduction; housekeeping; myosin-II.
XX
OS Homo sapiens.
XX
PN W0200218946-A2.
XX
PD 07-MAR-2002.
XX
PF 23-AUG-2001; 2001WO-GB03791.
XX
PR 23-AUG-2000; 2000GB-0020833.
XX
PR 15-DEC-2000; 2000GB-0030637.
XX
PA (BABR-) BABRAHAM INST.
XX (UYCA-) UNIV CAMBRIDGE TECH SERVICES LTD.
XX (LIMZ/) LIM Z.
XX
PI Stephens L, Hawkins PT, Holmes AB, Manifava M, Ktistakis N;
PI Thurling JWCJF;
XX
DR WPI; 2002-434908/46.
XX
DR N-PSDB; ABK52205.
XX
PT Novel probe for use in assay method for detecting, measuring,
PT identifying and/or isolating PA- and/or PIP-binding protein in a test
PS sample, has immobilised phosphatidyl acid attached to solid support -
PS
XX Disclosure; Page 134; 163pp: English.
XX
CC The present invention relates to a new probe comprising or consisting of
CC an immobilised phosphatidyl acid derivative attached on to a solid
CC support, or a phosphatidic acid (PA) functionalised solid support. The
CC probe of the invention is useful in an assay method for identifying
CC and/or isolating a protein that binds to the probe. The invention is also
CC useful for detecting, measuring, identifying and/or isolating more than
CC one type of phosphatidic acid and/or phosphoinositide-binding protein
CC (PIP) from a test sample e.g. a tissue of tissue culture extract. The
CC assay is also useful for detecting, measuring, identifying and/or
CC isolating phosphatidic acid and/or phosphoinositide-binding proteins in
CC a test sample, to detect and/or measure the ability of an agent, applied
CC to phosphatidic acid and/or phosphoinositide-binding protein-containing
CC test sample, to agonise or antagonise protein-probe binding, and to
CC detect and/or measure the ability of an agent, applied to the probe, to
CC agonise or antagonise protein-probe binding. The probe is useful for
CC identifying an agonist or antagonist of phosphatidic acid/
CC phosphoinositide-binding protein-phosphatidic acid/ phosphoinositide
CC interaction, and in a single step high throughput screen of candidate
CC agonist and/or antagonist. The invention is also useful for identifying
CC important proteins for signal transduction, housekeeping and diagnosis.
CC The probe is useful as an important research tool in fundamental research
CC for diagnostics and drug discovery. The present amino acid sequence
CC represents the human myosin-II protein MYOLF of the invention.
XX
SQ Sequence . 1098 AA;

```

```

Query Match      31.8%; Score 1714.5; DB 23; Length 1098;
Best Local Similarity 38.1%; Pred. No. 1.1e-136;
Matches 400; Conservative 152; Mismatches 304; Indels 193; Gaps 23;

```

```

QY 28 GVQDVLLENFTSEAFLENLRRLRRRENTIYTYIGPVLYSVNVPYRDLOIYSQIHMERRG 87
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 18 GVDVNLPLQIT-EDAIANLRKRFMDYITFYIGSVLYSVNPFKQMPYFTDEIDLQYG 76

```



QY	88	VSFEYEPHLEAVADTVYRALTERRODAVNIISGSSGKTEATKRLLOFAYETCAPER	147
Db	77	AAQVENPHRYALTDNMRNMLIDECNOCVIISGSGAKTVAAKYIMGYISKVSGGK	136
QY	148	GAVALNDRLOSPRYLEAFSGNKKTLYANDNSSFFGKYVMDOVEFKAPVGNHLSYLEKSR	207
Db	137	VOHVADILQSNPRLAEAFGNKFTVRNNNSSFFGKYFEIOTFSRGSEPDGKISNMLEKSR	196
QY	208	VVHONIGERNFHVYUOLLEGGEEYTLRLGLERNPOSYLYLVUGGQCAKSSINDKSDMKV	267
Db	197	VVMONENERNFNHYUOLLEGGASQOQRONGL- MPRDYTYLLNOSDHYUODGTDDBRDFOE	255
QY	268	MKRALSYIDFTEDEVEDLLSVASVHLGNIHNADEDSN- AQVTTENDOKYTLRLGVE	326
Db	256	TLSAMQYIGIPRSTQOLVLOLVAGILHGNISFC- EDGNYAVEVSDVLLAFPAVYLGLD	313
QY	327	CTTLREALLTRKITA- -KGBELSPRLNEQAARYDALAKAVYSRTETWLYRKINSL	382
Db	314	SGRIQDEKLYSRKMDSRWGSGSESIINTLVNEQAARYDALAFDLFEVLEAINRM	373
QY	383	ASKDASPSWRSPTYVLGLDIYGEFEVGNSTFORCINCNELQOLFTELTKEOEY	442
Db	374	-OKPOEYS- -IGVLDIYGFELFQANGEOFCINFVNEKLOQOIFTELTKEOEY	426
QY	443	EAGSIMEVVOYFNNKTIJODLVEERFK- GIISILDEEL- -RPGCATDLYTEKLEDT	497
Db	427	VOEGLIMPTIOYFNKKVUCDLIENTLSPRGISVLDVCAIMNATGGGADOTLLLOA	486
QY	498	VKNPHLEPLIKLADOKTRKSLDRGFEPLIHYAEGVTSYVGEJDKNDLLERLNETMCS	557
Db	487	VOGTHEHF- -NSMSAG- FVHNHAGKSYVDSVSGFCERNRYFSDLEIEMOT	534
QY	558	SNMPLMOCFOKSELSDDK- RPEVYANOQKMSLLOVELLESKEPAYITRCIKRPDADQRP	616
Db	535	SEQALRLRLFPKLDGDKGRPSFAGSKIKKQANDVATLMRCRPHIRICIKREKETHAR	594
QY	617	REDEVLYHNOYVYGLMNLNLYVRAGFYRKYKVAEFLQKSLCSETPYMAAGRPDQVA	676
Db	595	DMEERNVYHIOVEYLCKENINVRAGAYAKRQAFKLORYAILTPETWPRMRDDEQVO	654
QY	677	VLVNRLYKAPREYKMGKTKFIRPKTYLFEATEDSELYEROSLATAIOAARGFHNQKFL	736
Db	655	HLRAVNMERPDQOMGSTKVFKNPBSLFLLEEYREKRFDGAFATIOKM- -	704
QY	737	RVKRSAICIOSMWKGTJLGRKRAKKAQAOTIRLLRGFLIRISPCPENAFLOHVRS	796
Db	705	- -RHVAVVRKY- -	713
QY	797	FLNLNRQLPRNVLDTSWTPPRALREASBELLRELCOMKNWVKYCSISPEWKOOLQA	856
Db	714	- -EENKE- -	718
QY	857	VASEIRKAKKDNYPQSVPRLEFSTIRLSTEISPRVLOSIG- SEPTQAVPVYKYDRKQY	915
Db	719	EASNILLNKERRRNSINENFVGYDLGEE- RPELRQELGKKERYADFSDVTKYDR- FK	776
QY	916	PRPQRLLPSPVAVYIEODK- -VKQIDVANTLGIVSYSSLSDSLYFVHY	962
Db	777	PLKRDLLTPKCYVYIGREKKMKKGPEKQOYCEVLKAKKVIDIALRGVSLSTQDDFFLL-	834
QY	963	QMEDNKQGDVVLQSDHVIETLTKF- -ALSADVRNNININQSSITFA- -	1007
Db	835	-OED- -ADSFLESVFKEPEYSLCKCFEETRRPLRLPDLQFVKKEGW	884
QY	1008	GGPRGDGILIDPTSGSELLITAKKGNILAV	1036
Db	885	GGGGRSVTFPSRGEGDIAVLKVGGRJTLV	913

RESULT	10
AAU23128	
ID	AAU23128 standard; Protein; 909 AA.

XX	AC	AAU23128;	
XX	XX		
DT	XX	17-DEC-2001	(first entry)
DE	XX	Novel human enzyme polypeptide #214.	
KW	KW	Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;	
KW	KW	lignase; hyperproliferative disorder; immunodeficiency disorder;	
KW	KW	autoimmune disorder; neurological disorder; metabolic disorder;	
KW	KW	inflammatory disorder; cardiovascular disorder; reproductive disorder;	
KW	KW	blood-related disorder; infectious disorder; cytostatic; anti arthritic;	
XX	OS	nephrotropic; anticoagulant.	
XX	OS	Homo sapiens.	
XX	PN	WO200155301-A2.	
XX	PD	02-AUG-2001.	
PE	PE	17-JAN-2001; 2001WO-US01239.	
XX	XX		
XX	XX	31-JAN-2000; 2000US-0179065.	
PR	PR	04-FEB-2000; 2000US-0180638.	
PR	PR	24-FEB-2000; 2000US-0184664.	
PR	PR	02-MAR-2000; 2000US-0186350.	
PR	PR	16-MAR-2000; 2000US-0189874.	
PR	PR	17-MAR-2000; 2000US-0190076.	
PR	PR	18-APR-2000; 2000US-0198123.	
PR	PR	19-MAY-2000; 2000US-0205515.	
PR	PR	07-JUN-2000; 2000US-0209467.	
PR	PR	28-JUN-2000; 2000US-0214686.	
PR	PR	30-JUN-2000; 2000US-0215135.	
PR	PR	07-JUL-2000; 2000US-0216647.	
PR	PR	11-JUL-2000; 2000US-0216880.	
PR	PR	11-JUL-2000; 2000US-0217487.	
PR	PR	14-JUL-2000; 2000US-0217496.	
PR	PR	26-JUL-2000; 2000US-0218290.	
PR	PR	26-JUL-2000; 2000US-0220963.	
PR	PR	14-AUG-2000; 2000US-0220964.	
PR	PR	14-AUG-2000; 2000US-0224518.	
PR	PR	14-AUG-2000; 2000US-0224519.	
PR	PR	14-AUG-2000; 2000US-0225213.	
PR	PR	14-AUG-2000; 2000US-0225214.	
PR	PR	14-AUG-2000; 2000US-0225266.	
PR	PR	14-AUG-2000; 2000US-0225267.	
PR	PR	14-AUG-2000; 2000US-0225268.	
PR	PR	14-AUG-2000; 2000US-0225270.	
PR	PR	14-AUG-2000; 2000US-0225447.	
PR	PR	14-AUG-2000; 2000US-0225477.	
PR	PR	14-AUG-2000; 2000US-0225757.	
PR	PR	14-AUG-2000; 2000US-0225758.	
PR	PR	14-AUG-2000; 2000US-0225759.	
PR	PR	18-AUG-2000; 2000US-0226279.	
PR	PR	22-AUG-2000; 2000US-0226681.	
PR	PR	22-AUG-2000; 2000US-0226688.	
PR	PR	23-AUG-2000; 2000US-0227182.	
PR	PR	30-AUG-2000; 2000US-0227009.	
PR	PR	01-SEP-2000; 2000US-0228924.	
PR	PR	01-SEP-2000; 2000US-0229287.	
PR	PR	01-SEP-2000; 2000US-0229343.	
PR	PR	01-SEP-2000; 2000US-0229344.	
PR	PR	05-SEP-2000; 2000US-0229345.	
PR	PR	05-SEP-2000; 2000US-0229509.	
PR	PR	06-SEP-2000; 2000US-0230437.	
PR	PR	06-SEP-2000; 2000US-0230438.	
PR	PR	08-SEP-2000; 2000US-0231242.	
PR	PR	08-SEP-2000; 2000US-0231243.	
PR	PR	08-SEP-2000; 2000US-0231244.	
PR	PR	08-SEP-2000; 2000US-0231413.	
PR	PR	08-SEP-2000; 2000US-0231414.	
PR	PR	08-SEP-2000; 2000US-0233080.	
PR	PR	08-SEP-2000; 2000US-0233081.	

Pr	01-DEC-2000;	2000US-0250391.	
Pr	05-DEC-2000;	2000US-0251030.	
Pr	05-DEC-2000;	2000US-0251988.	
Pr	05-DEC-2000;	2000US-0256719.	
Pr	06-DEC-2000;	2000US-0251479.	
Pr	08-DEC-2000;	2000US-0251856.	
Pr	08-DEC-2000;	2000US-0251868.	
Pr	08-DEC-2000;	2000US-0251869.	
Pr	08-DEC-2000;	2000US-0251989.	
Pr	08-DEC-2000;	2000US-0251990.	
Pr	11-DEC-2000;	2000US-0254097.	
Pr	05-JAN-2001;	2001US-0259678.	
XX			
PA	(HUMA-)	HUMAN GENOME SCI INC.	
XX			
PI	Rosen CA,	Barash SC,	Ruben SM;
DR	WPI: 2001-465566/50.		
XX	N-PSDB: AAS40998.		
Pr	Novel polypeptides and polynucleotides useful for diagnosing, preventing, treating neural, immune system, muscular, reproductive, pulmonary, cardiovascular, renal, proliferative disorders and cancerous diseases		
PT	-		
PS	Claim 11; SEQ ID No 1124; 1180bp; English.		
XX			
CC	The present invention relates to the isolation of novel human enzyme polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences encoding them. The enzyme polypeptides of the invention may comprise the functional classes of oxidoreductases, transferases, hydrolases, lyases, isomerases or ligases. The sequences of the invention are useful in the diagnosis, treatment, prevention and/or prognosis of a wide range of disorders, including hyperproliferative disorders (e.g. cancer), immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis), blood-related disorders (e.g. haemophilia), reproductive disorders (e.g. infertility) and infectious disorders (e.g. influenza). The polynucleotides of the invention can also be used in gene therapy.		
CC	AAU22915-AU23814 represent the novel human enzyme polypeptides of the invention.		
CC	Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct-sequences.		
CC			
XX			
SQ	Sequence	909 AA;	
Query Match			
	Best Local Similarity	31.2%;	Pred. 1685; DB 22; Length 909;
	Matches 376; Conservative 133; Mismatches 292; Indels 60; Gaps 18;		
QY	28	GVQDFVLENTSEAFENTLRFRFRLTYTIGPVLSVNPYRDLOYSRQHMERYG	87
	I		
DB	18	GRPFDVLQDVYME-DEMRLQLRFEKGRITTYIGELVSNPYQELPLYPEALARQG	76
QY	88	VSEFVEYRPHLFAVADTVYRRLRTERRDQAVMIGSGSAGTTEATKRLLOFYAETCPAPER	147
	I		
DB	77	RELXYERPHLYAANAAYKAMKRRSRPTCIVISGSGAGTTEASKHIMQYIAAVTNPFSOR	136
QY	148	GGA--RDRILQSNPYLEARGNAKTLINDNDSRRGKTMDOQDFDKGAPVGHILSYLLEK	205
	I		
DB	137	AEYERKADLLSTCYLEARGNARTNNHSSRGKTYMDINFDKGPDIIGHIISYLLEK	196
QY	206	SRVYHONHGERNFHVEYOLLDEGGEETLRRLGLERNQSYLYLVKGCACAKY---SSI-ND	261
	I		
DB	197	KRYLKHQVBERNFHAYTOLRGSSEDKQLHLLHLEARNVAVNFTHQAGGLNMTVTSALDSD	256
QY	262	KSDMKVMKRALSYIDTFEDEVDLSTIVASVLIHGINHPAEDDSNAO---VTENQL	316
	I		
DB	257	EOSHQAVVTAMRVIIGSPSEVESVYHRIILAIILHGINIEFETEEBGGLQEGILAAVEAEALV	316

Qy	317	KLITRLLLGVEGTTILREALTLHKRIIAKGEELLSP-LNLEQAAVARDALAKAVYSRTFTVLV	375
Db	317	DHVAELTAPPRDLVLRSLKSLARTVASGGRELLIEKHTMAEASVADACKAKAYQLEFEMV	376
Qy	376	RKINRSASKDAESPSWRSSTVTLVGLLDLYGFEVFNQHSFEOFCINYNENKIQOLEFELTL	435
Db	377	NRINSEMERGRDPRRDKDGYISGLDLYGFEVFNVSFEQFCINYNENKIQOLEFELTL	436
Qy	436	KSEQEEVEEGIAWEPVOYFNKKIICDLVEEKFGKIIISLDEECILRGEGATDLTFLEKLE	495
Db	437	KQEOEEVEEGITWQSVSEYFNNAITVDLVERPHGILAVLDEACSSAGITTDKILFQTLD	496
Qy	496	DTVPRPHPLTHKLADQTRKSLRG-EFRLHYAGSVTYSVGLDKNNDLFRNLKET	554
Db	497	THNRHNLHYTSRQLC-PTDQTMEEGDFRIKHACGVTVSVGEFDKNDFLEDFQFRLL	554
Qy	555	MCSSNNPIMACF-----DKSELSDKKRPETVATOFKMSLLQVELLRSEKPAVIRCIKP	609
Db	555	LYNSTDPLLRAMWDGQDDITEVY--KRPITLGTGLEKNSVWLVLENLASKPEPTVRCIKP	612
Qy	610	NDAKOPRFEVLIRIQVKTGLGEMENLRVRGFAIRKRYEAFLOKYSLCPETWP-MMA	668
Db	613	NEDVAKGLDENICRHQVAVLGLLENNVRARRGFASROPYSFRLRYKMTCEYTPNHLL	672
Qy	669	GRPDGAVLVRLHLYGKPEYKMKGRKIFIRPKTIFATEDSLLEVROSLATKIQAAARG	728
Db	673	GSQKAVSALDEHGLQ-DDVAFGSHSLFIRSPRLVLTLEQ-----	713
Qy	729	FHKRQKFLRYKRSALICIQSMWGTLCGRKAARKMAAQTIRLLRGFLRHSPPCPENAF	788
Db	714	--KARLLIP--IVLLQKAMRGTLARRR-CHRLAITYITMRW---RIKVARA-----	758
Qy	789	FLDHYVASFLLNRQLPRNVLDTSWTPPRPALREASSELLRELCKMNMWYKCSISPEW	848
Db	759	HLAEIQRR--QAAPQPLYGKDLWPLRPVAVLQPFQDTHALFCRRARQVLXKNIPSD	816
Qy	849	KQQLQOKAVASIEFGKKDNY 869	
Db	817	MPQIKAKVAAAGALQGLQDQW 837	
RESULT 11			
ABB97258			
ID	ABB97258	standard; Protein; 1089 AA.	
AC	ABB97258;		
XX			
.DT	27-JUN-2002	(first entry)	
XX			
De		Novel human protein SEQ ID NO: 526.	
XX			
KW		Human; anti-naemic; vulnerary; antiinflammatory; immunomodulator;	
KW		antinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy	
KW		neuroprotective; antiparkinsonian; protein therapy; EST;	
KW		expressed sequence tag.	
OS		Homo sapiens.	
XX			
PN	WO200222660-A2.		
XX			
PD	21-MAR-2002.		
XX			
PR	10-SEP-2001; 2001WO-US26015.		
XX			
PR	11-SEP-2000; 2000US-0659671.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;		
XX	Xue AJ, Yang Y, Wehrman T, Dimañac RT;		
XX	WPI; 2002-292408/33.		
DR	N-PSDB; ABN32444.		

XX An isolated polynucleotide for treating diseases associated with its  
PT encoded polypeptide such as cancer and multiple sclerosis -  
XX  
XX  
PS Example 2: SEQ ID NO 526; 509pp; English.  
XX  
XX  
CC The present invention provides the protein and coding sequences of 444  
CC novel human proteins. These were isolated from expressed sequences tags  
CC (ESTs). They can be used to stimulate cell growth, to regulate  
CC haematopoiesis e.g. to treat aplastic anaemia, to regulate  
CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat  
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat  
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions  
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.  
XX Parkinson's disease. The present sequence is a protein of the invention.  
XX  
SQ Sequence 1089 AA;

Query Match	29.0%;	Score 1565;	DB 23;	Length 1089;
Best Local Similarity	35.7%;	Pred. No. 6.3e-124;		
Matches 378;	Conservative 147;	Mismatches 232;		

70 РYBPIATYCHOV, ZHURAVLEV 203; Indels 252; Gaps 25;

Db 2 PY-----FTDREIDLYOGAAOVENDUWVITRRAALTEERRDOAVMISGESGACK

...KQSSDYLQSSAQLENPFLYALFDNMYRNMLIDCENQCVIISGESGAGKT

130 ATKRLLOFYAETCPAPERGGAVRDLLQSNPVL EAFGNAKTLRDNSSREGKYMDOFT

Db 57 AAKYIMGYISKVSGGGEKVOHVKDITIOSNDITETACGVAATTTT

00 190 KQADPUCUWZ 0000

150 AGAPVGGHLLSYLLEKSRVHQNNGERNFHFYOL-----

Db 117 GGE PDGKISN FLL EKS RVV MONE NERN EHTYVOCCPDCGCGC

0V 225 .  
-----  
ZNYVQGNAGGEGAPPGAQDRGGQWEGKHL

-----LEGGEEETLRLGLERNPQSYLYLVKQOCAR

Db 177 GHGRADEGSLQASTRCMMATPSSWLEGAQFOFOPICT - MEDIAN

OV 257 SSINDKSDKRVNDVAI SUTR-----  
-----EONQEQVNEGL-MIPDYXXYLNQSDTYQ

SYNDKSNVWKALSVIDEVEDLSIVASVLHGNHFAEDSN-AQVTTENC

Ddb 236 DGTDDRSEGETLSAMQVIGISIOOL--VIOIVAGTIHICNTEEC--

316 I KYITBI I CBECEMT BHTTHTT

-----KGEELSPLNLEQAAAYARDALAKAVYSRTE

292 LAFPAYLLGIDSGRLQEKLSRKMDSRWGGRRSESTNYVTI NYEQAYEDPT

372 TWIYRKINBCTACVCD,DCGCTT  
RDALAKGLYARLE

[illegible]

352 DFLVEAINRAM-QKPQEYS-----IGVLDIYGFEIFQKNGEEOCTINEUNETCOIT

432 EITIKSEOFYEAECTATTPDUC.....

```

|||||:|||||:|||||ICDLVEEKFK--GIISLDECL--RPGEAT

```

405 ELTLKAEQEYVQEGIRWTPIQYFNKVVCDLIENKLSPPGIMSVLDVYCATWUWTCG:

487 DLTFLFKL.FDTVBPUH MUK \*C\*\*\*\*\*  
.....LDDVCAIMHAI GGA

[illegible]

465 DQTLQKLQAAVGTHEHF-----NSWSAG-FVITHYAGKVSVDVSGCEFPBNDV

Y 547 LERNKETMCSMBPTMAOCEDKCEI CCCC

REPEVATQFKMSLLQVLEILRSKEPAYIR

513 LFSDLIELMQSSDQAFRLRLFPEKLDGDKKGRPSTAGSKIKKQANDLVATIMRCTPHYTP

606 СИРНАКОПРЕДЕЛ.ТРИОВКИ СІМЕН ВУДІ...

[illegible]

3/3 C L K P N E T K R P R D W E E N R V K H Q V E Y L G L K E N I R V R A G F A Y R R Q F A K F L O R Y A I T P E T W P

666 MWAGRPQDGVAVLVRLHGYKPFYKMGPTETDDNKCT

[illegible]

055 KMRGDERÜGVÖHLKRAVNMEPDÖYQMGSTKVFKNPESLFLLEEVREKKEFDGEARTIOKA

7 26 WRGEHWROKFLRVKRSACICIOSWWRGTIGPRKAKBRYAJOETNN

Db 693 W-----RRHVAVRY-----702  
Qy 766 NAFELDHVASEFLNLRRQLPRNVLDTSWTPPPALREASELLRELCMKNMWYKCRSIS 845  
Db 703 -----EEMRE-----707  
Qy 846 PEMKQOOLOKAVASEIFKGGKNDYPSVPRLEFSTRLGTEESIPVLSLG-SEPIQYAV 904  
Db 708 -----EASNILLKERRRNSINRNFVGYLGE-EPELRQFLGKKEKREVFAD 755  
Qy 905 PVVKDRKGYPRRQLLTPSAVIVEDAK-----VKORDVANLTGISVS 951  
Db 756 SVTKYDRR-FRPIKRDLLTPKCYVIGREKKMKGPCKPVCCELLKRLDIALRGVSLS 814  
Qy 952 SLSDSLFVLHVGREDNKKQGDVVLQSDHVIETLTKT---ALSADRVNNINIGSITFA- 1007  
Db 815 TRQDFEFL---QED-----AADSFLSEVFTEFVSLCKRFEETRRPLFTFS 862  
Qy 1008 -----GGPRDGIIDFTSGSELLTPKAKNGHLAV 1036  
Db 863 TLQFHVKKKGCGGTSRVTSRFGDLAVLKVGGRITLV 902

RESULT 12

ABG20610 ID ABG20610 standard; Protein: 759 AA.

AC ABG20610;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #20601.

KM Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KV food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HSE-) HYSEO INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR N-PSDB: AAS84797.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

PS Claim 20; SEQ ID No 50969; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (II) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPD  
at ftp.wipo.int/pub/published\_pct\_sequences.

SO Sequence 759 AA:

Query Match 26.2%; Score 1412.5; DB 22; Length 759;

Best Local Similarity 38.5%; Pred. No. 3.7e-111;

Matches 332; Conservative 122; Mismatches 261; Indels 147; Gaps 18;

Qy 28 GYODVVLLENFTSEAFLENLRRENNILTYIGPVLYSVPRDLQIYSQOMHERYRG 87  
Db 10 GPRDEVLLDQVTME-DVFNILQFREKGRITYIGEVLSVNPYDELPLYGEALARYOG 68  
Qy 88 VSEFEVPRPLFAVADPVYVYALRTERDQAVMTSGSGAGCTATKRLQFYAETCPAPER 147  
Db 69 RELTERPRPLVAVANAVYKAMKRRSDCTIVTSGESAGKTASKIMQYIAAVNTPSOR 128  
Qy 148 GGA--VRDRLQSNVYLEAFGNAKTLRNDSSRFKYMVDPFKGAPEVGHILSYLTK 205  
Db 129 AEVERKQDVLLKSTCVLEAFGNARTNRNNSRFKGYMDINDFKGDPIGHIHSYLTLEK 188  
Qy 206 SRVAVQNGEGRNFHYVQYLLGEGEETLRRLGLERPNQSYLYLVKGQCAKVSINDKSM 265  
Db 189 SRYLKQHGVERNFHAFYQALD-----SDEQSH 215  
Qy 266 KVMRKALSYIDFTEDEVEDLLSIVASVHLGNITHPAEDSDNAO-----VTTEENQLYLT 320  
Db 216 QAVTEAMRYIGSPREVEVHRIIAIILGNIETVEEGLOEGLAVAEALVDHVA 275  
Qy 321 RLLGVEGTTLRALTHRKLIKAGEELLSP-LNLEQAAVARADALAAVYSRTFTWLKRN 379  
Db 276 ELTATPRDLVLSLLARFVASSGRELIEKGHTAAEASVARDACAKAVYQRLPEWVYNRIN 335  
Qy 380 RSLASKDAEFSWRSTVYGLDIYGFVQHNSPQFCINCKEQLQOLFIELTLSQ 439  
Db 336 SVMEGRGRDPRRDGDTYGVLDIYGFVQHNSEFCINCKEQLQOLFQTLKQ 395  
Qy 440 EYEBAEGLAMEPVQYFNKNIICDLVEKFKGIIITIDECLPGEATDFTPEKLEDPVK 499  
Db 396 EYEREGITWQSVETFNNAFTYDLVERPHRGLAVLDEACSSAGITTDITFTQIDMHR 455  
Qy 500 PAPHLETHK-----LADQKTRKSLDRG-EFRLLHYAGEVTYSVYGLDKNNDLFRNLK 552  
Db 456 HHLHTSRQVPVAVVPPQWADKTMEFGRDPRFKHAGD-----DLK 497  
Qy 553 ETWCSSMNPIMACF-----DKSELSDKRPETVATQFMSLQVLEILRSKPAYIRKI 607  
Db 498 RMLYNSTDTPLKAMPDGOODITVEYV--KRPLTAGLTFKNSWALVENLASKEPEYVRCI 555  
Qy 608 KPNDAKQPRFPEVLIRHGVKYLGMENLVRRAGAYRKTKEAFLQKSKLCPETWPMW 667  
Db 556 KPNEDKVGKLDENHCRRQVATYGLLENAVS----- 587  
Qy 668 AGRPDQVAVLVRLHLYGKPEEYKMGRTKIFTFEPKTIFFATEDSLLEVROSATKIOAMR 727  
Db 588 ALKEQHGIL-----QGDVAFGHSKFIISPRLVLEGS----- 620  
Qy 728 GFWNRQKFLRVKRSATCIOSWNRGTLGRKAARKKAAOTIRLLIRGTLRHSPPCEPA 787  
Db 621 ---RARLPI--IYLLQKAMRGTLARWR-CRRRLAITYIKRWF-----RKVRA----- 665  
Qy 788 FFLDHYRASEFLNLRRQLPRNVLDTSWTPPPALREASELLRELCMKNMWYKCRSISPE 847  
Db 666 -HLAEIQRFF--QAAKQPLPYGRDLWPLPRAVLQFPQDTCHALFCMKRAROLVKNIPPS 722  
Qy 848 WKQOOLQOKAVASEIFKGGKNDY 869

Db 723 DMPQIKAKVAMGALQCLRODM 744

# RESULT 13

ABR60369 standard; Protein: 2167 AA.

XX ABR60369;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 7899.

XX Drosophila: developmental biology; cell signalling; insecticide; pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW.

XX WPI: 2001-656860/75.

XX N-PSDB: ABL04472.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

XX Disclosure: SEQ ID NO 7899; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB101840-AB101875), expressed DNA sequences (AB857737-AB857742) and the encoded proteins

XX The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 2167 AA:

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Db 211 FGKYIDHFSANGVIEGAKIEQYLEKSRIVSONHSEFNHVFICILAGLSADEKSRDL 270

Qy 237 GLEHNPQSYLYLVKGQCKAVSSINDKSDWKYMRKALSIDFTEDEBDLSIYASVLIHQ 296

Db 271 GMAAD---YKLTGNSITCEGRDDAEFSDIRSAMKVLLEFSQEIWEIETIKLALLHGG 327

Qy 297 NIHFPA--DEDSNAQVTTENQULYTRFLGVEGTTREALTRKTIANGCELLSLPNE 353

Db 328 NIKKATVVDNLDATEIPEHINVERVAGLGLPLPPLIDALTRTLPAHGETVSTLSRD 387

Qy 354 QAAVARDAKAVYSKRTFTLVKRNKSLASKAESPSMSTTVGLDIYGEFVQHS 413

Db 388 QSDVDRDAFEWKGIGYGRMEFVHIVRKINTAIF----KPRGTSRAIGVLDIFGEFNDMS 442

Qy 414 FEQFCINVCNEKIQQLFETLTKSEOEVEAEGIAEPQYFNKTKICLVEEKRGITIS 473

Db 443 FEQFCINVCNENIQQFVQHIIFLEQDETNHEATINQHIQFVNDQDALDLIAIKQINIMA 502

Qy 474 ILDEECURGEATDITFLEKLEDTYKPHHFLTHKLADQTKRSLDRGFRLHYAGEVT 533

Db 503 LIDDEARFP-KGTDQTMIAKHTHSHKNYLKP-SDINT-----SFGLNHFAGYVF 553

Qy 534 YSVTGFLDKNNDLLEFNLEKTMSSMNPMAQCFLKS---ELSDKKRETVATQFKMSL 590

Db 554 YDTRGFLDKNNDLTFSPDLHLVQSSTNKFRLQIFADIDEMGAETRRKPTLSTQFRKSLD 613

Qy 591 QLVELRSKEPAVIRICIKPNDAKQGRFDEVLRHGVKYLIGEMLRVRRGEAFARKYE 650

Db 614 ALMKTLSSCOPEFIRICIKPNELKPMFDRGLCCQQLKSGMETIRTRRGVPRIGER 673

Qy 651 AFLQKSLCEPEWPMWAGRPDQVA-VLVHRLGKPEEYKMGRTKIFIRPKTLFATED 709

Db 674 EYVERTRFLIPVPAHRTDCAATSRIQAVLQ--KSDYDGLGHTKVPILKDAHFL---- 727

Qy 710 SLEVRQSLATK---IOAMRGFMHROKFLRKVSATCISQSMRGTIGRKAARMAA 765

Db 728 -LEQERDRLVTRKILILQRSIRGWYRRRFLRLRAALITVQRFWKGIAQRKRRNMAGV 786

Qy 766 QTRRLRIGETIRH 779

Db 787 MRLQALIRSVLSH 800

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

```
XX 31-MAR-2000; 2000WO-US08621.
PF
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
PI
PI Shinkets RA, Leach M;
DR
DR WPI: 2000-602362/57.
DR N-PSDB; AAC76860.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
PT
PS Claim 11; Page 4015-4016; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antiproliferative; antiparinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiallergic; immunosuppressive;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatologic; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antilyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
CC
XX
XX Sequence 512 AA:
SQ
Query Match 24.6%; Score 1330.5; DB 21; Length 512;
Best Local Similarity 53.9%; Pred. No. 2e-104;
Matches 260; Conservative 82; Mismatches 129; Indels 11; Gaps 4;
QY 27 VGVDFVLENTSEAFIENLRFRRENLIYTYIGPVLVSNPYRDLQIYSRQHMERYL 86
DB 15 IGVDFMVLLEPL-NEFTINMLKRFHSEIITYIGSVIVSNPRSLPITSPEVEERY 73
QY 87 GVSFEVPPHLEFAVDIVYRALFTRERDQAVMISGSGAGTEATKRLQLQYATCPAPE 146
DB 74 NRNFELSPHIFALSDDEAYRSLRDKDQCILITGSEAGKTEASKLVMSVAACGKA 133
QY 147 RGAVRDLLOSNPVLEAFNAKTIARNSSRFKGMVGVODPKAPVCGHLSLLEKS 206
DB 134 EVNOVKEDLOSNPVLEAFNAKTIARNSSRFKGMVGVODPKAPVCGHLSLLEKS 193
QY 207 RVVHONHGERNFHFVYOLLEGGSEETLRLGLERNPOSYLVLVKGQCAKVSINCKSMK 266
DB 194 RVVQOPRGERNFHFVYOLLSGASEELNKLKLERDFSRNYL-SLDSKAKVGVDAANFR 252
QY 267 VMKRALSVIDFTEDEVEDLSTIVASVHLGNIFPAADEDN---AQTTEQNLKYLTRL 322
DB 253 TVRNAMQVLFPMDEHAEFVLAIVAVALKLTGNIIEFKPESRVANGLDISKIKDKNELKEICEL 312
QY 323 LGVEBTTIREALTHKRIKAKGEBLLSPNLBQAVARALAKAVSRFTWLVKRIINSL 382
DB 313 TGIOSVLERAFSFTVEAKQEKVSTLTINVAQAVYARADALAKNLVSRLEFVWLNRINESI 372
```

```
QY 383 ASKDAESPWSSTVVLGLDIYGFVFOHNSFEOFCINVCNEKILQOLEITELTKSEOEY 442
DB 373 KAOTKVRKK-----VMGYLDIYGFIEFDNSPEOFTINVCNEKILQOIFTELTKSEOEY 427
QY 443 EAEGIAMPEPVQYFNNKTIICDLVEEKFGIISLDECLRPFGATDLTLEKLEEDYKRRP 502
DB 428 IREDIEWTHIDYFNNAIICDLIENTNGIAMLDECLRPFGVTDTELEKINOVCAHQ 487
QY 503 HF 504
DB 488 HF 489
RESULT 15
ABB62828
ID ABB62828 standard; Protein: 2129 AA.
AC
AC ABB62828;
DT
DT 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 15276.
DE
DE Drosophila: developmental biology; cell signalling; insecticide;
KW pharmaceutical.
KW Drosophila melanogaster.
OS
OS WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
PF 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
PA
PA Venter JC, Adams M, Li PWD, Myers EW;
PI
PI WPI: 2001-656860/75.
DR
DR N-PSDB; ABL06931.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
PT
PT Disclousure; SEQ ID NO 15276; 21pp + Sequence Listing; English.
PS
PS
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL101840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
SQ
SQ Sequence 2129 AA:
Query Match 24.6%; Score 1328; DB 22; Length 2129;
Best Local Similarity 34.7%; Pred. No. 3e-103;
Matches 329; Conservative 165; Mismatches 331; Indels 122; Gaps 22;
QY 29 VODFVLENTSEAFIENLRFRRENLIYTYIGPVLVSNPYRDLQIYSRQHMERYRCV 88
DB 74 VEDMTTLDL-QEYTIIRLQNRVAKOLITYTGSMLVAINPYOILPVTNREIDLRYRK 132
QY 89 SFYEVPPHLEFAVDIVYRALFTRERDQAVMISGSGAGTEATKRLQLQYATCPAPE 148
```

Db 133 SLAELPHIEAISDNARFORLORLEKENCOCVVISGEGAGKTESTKLILQYLA---AISGH 189  
QY 149 GAVDRLLQSPVLEAFGNKAKTLANDSSRFGKMDYQFDEKGPVCGHILSTYLEKSRV 208  
Db 190 SWIQOILEAPLEAFGNKATVRDSSRFGKYLEIRFTPOGAIQARIQOYLLKSRRI 249  
QY 209 VIONHIGERNHFVQLEEGEETLRRLGL-ERNPOSYLVLVKQCAKASSINDKSDKV 267  
Db 250 VFOGRDERNRHIEFCMLAGLSTARERKLQEQSPSOYHILAQGCFTLPGRGDAKDFAD 309  
QY 268 MKRALSYIDFTEDEVEDLSIVASVHLGNIHFADEDSN---AQVTTENQLKYLTRLG 324  
Db 310 IRAMKVLSEKPEEWSILSLAAILHGNLRFATEVANLATAEIDTPTNLQRAQLLG 369  
QY 325 VEGTTLREALTHRKILAKGELSPLNLEQAAVARADALAKAVYSRTFTMLVKRINKSLAS 384  
Db 370 IPTSALNALTKORTIFVHGEHVTTSLSKENAIEGRODAFVKSLYDGIPIVIRIRINETI-N 428  
QY 385 KDAEBSRSTVYGLADIYGFVFOHNSPEOFCINYNKELQOLFELTLKSEOEYEA 444  
Db 429 KOYDOP---MNSIGYLDIFGFENFDNNSFEQCIYANENLQOFFVGHIFKMEODEYON 484  
QY 445 EGIAHEPVOYENKICDLYEKKGIISTLDEBCLRPGATDITLEKLEDTVKRPHHF 504  
Db 485 EHNMQHIEQDNOQILDLGKPMNLMSLIDESKFP-KGTQDTLEKLEH----- 534  
QY 505 LTH-----KLADQKTRKSLDRGFEFLHVAEYTVSGFLDKNNDLFPNLKETWCSSM 559  
Db 535 VOHGNRSIYKKGTTQTSL---FGIRHAYGVMYNPLGLEKNRDSFSGDLTKLVORST 590  
QY 560 NPTMAQCFDKSELSL---KKPEYVATQFKMSLQLOVEILRSKEPAYIRCIKPNDAKOPGR 617  
Db 591 NKLYVDIFPHEMDMDHAKOP-TLCVKFRNSLMLMRTLQAHPIFRICIKPNEYKEPKN 649  
QY 618 FDEVLIRHOVKYGLMENLRVRBAGFAYRKYBAFLQRYKSLCPETPMMAGRPOGVAV 677  
Db 650 FDKELCVROLRYSGMETARIRAGTPIRHAYRAFAVERYRLLPVGPPL---ECCDCRK 705  
QY 678 LVNHLGY---KPEYKMGRTKIFIRFPKTLFATED--SLEVRQSLATK---IQAMWR 727  
Db 706 LAHOICEVALPADSDROGYKTKLFLR-----DEDDASLELQORSQMLKSTVTLQIR 758  
QY 728 GFHWROKFLRVKRSALCISQMMRGTLGRK-----NAKRMAAQTTIR-- 770  
Db 759 RVLFRRYMKRYREALITVORYWGRGLQRRKYOMROGFHRLGACIAAQULTTKFTVRCR 818  
QY 771 -----LIRGFILRHSPRCPENAFFLDHRASFLNLRQLPRNVLDTSMTPTPALREA 824  
Db 819 TIKLOALSGYL-----VRKDFOKKLLERKONOLKKEELKLAKMKKEA 862  
QY 825 SELLELCKNNWVKYCRSISPEMKOOLQOKAVASEIFGKKKNYPQSVYRLFISTRIGT 884  
Db 863 EELLR-----LQOLKEOK-----EROREQOEKRLQEQRLKA 895  
QY 885 EETSPVLQSLGSEPIQYAVPVVKYDRKGYKRPQOLLTTPSAVITV 931  
Db 896 EAAARNALMAVAVOOKRRTKPVKQEPKAPVTLQARNSLCPPPTTLIV 942

Search completed: December 9, 2002, 16:44:13  
Job time : 73 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2002, 16:44:21 ; Search time 26 Seconds  
(without alignments)  
1181.443 Million cell updates/sec

Title: US-09-893-371-1

Perfect score: 5398  
Sequence: 1 MRYRASALGSDGVRTMESA.....LITKAKNGHLAVVAPRLNSR 1044

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A-COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B-COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A-COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B-COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCUS-COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1191	22.1	1939	US-09-310-187A-1	Sequence 1, Appli
2	1146	21.2	1120	US-09-147-404-1	Sequence 1, Appli
3	1105	20.5	2548	US-09-172-422-1	Sequence 1, Appli
4	1018	18.9	1886	US-08-938-105-3	Sequence 12, Appli
5	160	3.0	1809	US-09-012-515A-12	Sequence 12, Appli
6	160	3.0	1809	US-08-360-144A-12	Sequence 12, Appli
7	160	3.0	1809	US-09-012-504A-12	Sequence 12, Appli
8	160	3.0	2549	US-08-471-112A-3	Sequence 12, Appli
9	160	3.0	2549	PCT-US95-06722-12	Sequence 12, Appli
10	152.5	2.8	103	US-08-905-223-395	Sequence 395, App
11	130	2.4	1529	US-09-134-001C-3945	Sequence 3945, Ap
12	126	2.3	567	US-09-134-001C-3762	Sequence 3762, Ap
13	124	2.3	1093	PCT-US93-03077-1	Sequence 1, Appli
14	123.5	2.3	1333	US-09-356-952-2	Sequence 2, Appli
15	122.5	2.3	1319	US-08-290-731C-2	Sequence 2, Appli
16	122.5	2.3	1336	US-08-290-731C-6	Sequence 6, Appli
17	122	2.3	652	US-09-438-833-5	Sequence 5, Appli
18	122	2.3	813	US-09-438-833-12	Sequence 12, Appli
19	122	2.3	826	US-08-785-241-6	Sequence 6, Appli
20	122	2.3	826	US-08-480-473B-2	Sequence 2, Appli
21	122	2.3	826	US-09-915-213-2	Sequence 2, Appli
22	122	2.3	826	US-09-148-547-2	Sequence 2, Appli
23	122	2.3	826	US-09-235-217-2	Sequence 2, Appli
24	122	2.3	826	US-09-380-662-23	Sequence 23, Appli
25	122	2.3	826	US-09-438-833-1	Sequence 1, Appli
26	122	2.3	826	PCT-US96-10251-2	Sequence 2, Appli
27	122	2.3	1657	US-08-287-959-1	Sequence 1, Appli

28	120.5	2.2	874	US-08-804-439A-15	Sequence 15, Appli
29	120.5	2.2	874	US-08-720-229-15	Sequence 15, Appli
30	120	2.2	2154	US-08-841-349-4	Sequence 4, Appli
31	117.5	2.2	1151	US-09-134-001C-3242	Sequence 3242, Ap
32	115	2.1	1200	US-08-840-006-5	Sequence 5, Appli
33	113	2.1	1151	US-08-840-006-6	Sequence 6, Appli
34	112	2.1	810	US-08-785-241-7	Sequence 7, Appli
35	111.5	2.1	1181	US-08-488-940-2	Sequence 2, Appli
36	111	2.1	1066	US-09-541-782-8	Sequence 8, Appli
37	111	2.1	1066	US-09-723-820-8	Sequence 8, Appli
38	111	2.1	1194	US-08-488-940-1	Sequence 1, Appli
39	110	2.0	787	US-08-574-763-2	Sequence 2, Appli
40	110	2.0	1786	US-08-973-462-8	Sequence 8, Appli
41	108	2.0	805	US-08-480-473B-4	Sequence 4, Appli
42	108	2.0	805	US-08-915-213-4	Sequence 4, Appli
43	108	2.0	805	US-09-235-217-4	Sequence 4, Appli
44	108	2.0	805	PCT-US96-10251-4	Sequence 4, Appli
45	108	2.0	1280	US-08-583-276-19	Sequence 19, Appli

ALIGNMENTS

RESULT 1					
US-09-310-187A-1					
Sequence 1, Application US/09310187A					
Patent No. 6358751					
GENERAL INFORMATION:					
APPLICANT: Benichou, Gilles					
TITLE OF INVENTION: Pedoseyeva, Eugenia					
TITLE OF INVENTION: Involvement of Autoantigens in Cardiac					
FILE REFERENCE: USF-090					
CURRENT APPLICATION NUMBER: US/09/310,187A					
CURRENT FILING DATE: 1999-05-12					
NUMBER OF SEQ ID NOS: 3					
SOFTWARE: FASTSEQ for Windows Version 4.0					
SEQ ID NO 1					
LENGTH: 1939					
TYPE: PRT					
ORGANISM: Homo sapiens					
US-09-310-187A-1					
Query Match					
Best Local Similarity 22.1%; Score 1191; DB 4; Length 1939;					
Matches 281; Conservative 142; Mismatches 303; Indels 64; Gaps 18;					
OY	4	RASALGSDGVRTM-----SALTARDRV-----GVDFVLLNFTSEAFIENLR	49		
DB	48	KAKILSREGKVIATENGKTVTVKEDVQLQNPFPKIDQMAML-TFLHEPAVLFNLIK	106		
OY	50	RREFENLITTYGPAVLSVNPYRDLOIYSROMHERRYGVSEYVPPHLEFVADTVYRALR	109		
DB	107	ERYAAMMITYVSGLCVYVNPYKMLPYVNAEVAAYRGRKRSEAPPHIFSIDNAYQYML	166		
OY	110	TERROAVMYTISESGAGKTEAKTRKLOFYAEFCPAPEB-----GAVDRILQSNP	160		
DB	167	TDREMOSTILITTESGAGKVTNTRVIOYFASIAIGDRKKMANANKTLEDOTIQAMP	226		
OY	161	VLEAFGNKTLTNDNSRREGKTYMDQFPGAPVGGHILSYLLEKSRVHQHNGERNFHV	220		
DB	227	ALEAFGNKTYVNDNSRREGKTFIRHFGATGKLASDIETVLEKSRVIFOLKAERNYHI	286		
OY	221	FVQLLEGGEEFLRLKGLGERNPQSTLYLVKGCACAVSSINDKMDKVMRKALSIDFTED	280		
DB	287	FVQILSNKPELIDMLLVITNPNPYDAFVSQGE-VSVASIDSEELMARDSDADVIGFTSE	345		
OY	281	EVEDLSTVAVLHGNIFHADE-DSNAOVTTENOLKTLTRLGVEGTTLEALTRHRI	339		
DB	346	EVAGYKLTGALIMHGNMKFKOKREDAEPDGTEDADKSAITLMGLNSDLKGLCHPRV	405		
OY	340	IAKGELLSPNLLEQAAVARDALAAVYSRFTWLVKRIINRSKADAEPSMRSTTVLG	399		





Db 167 TDREKQSLITIGESAGCTVNTKRIQYFAYIAAIGDRSKDQDSGKCTLEDDIIOANPA 226  
Qy 162 LEAFNMAKTLRNDSSRRGKYMDVOFPDKGAPVGCHILSYLLEKSRVYHONHGERNFHV 221  
Db 227 LEAGNATVANDNSSRRGKFIHFGATXKLASADIEFYLEKSRVIFQKAEADYHIF 286  
Qy 222 YOLLEGEETLRRLGLERNPOSYLVLVGKQCAKVSINDSKVMKRAISVIDFTEDE 281  
Db 287 YOLISNKKRELLDMLLTNNPNYDAFISQGE-TTVAISIDAEELMATONAFDVLGFTSEE 345  
Qy 282 VEDLLISVAVLHGNHIFADE-DSNAQVTTENQKLKTLRLGVEGTLREALTHRKII 340  
Db 346 KNSMKRLGALMHFGNMFKLQREDOAPDGTEDRKSAYLIMGINSADLLKGLCHPYVK 405  
Qy 341 AKGEELLSPNLDEQAAYRADALAKAVYSRTFTWYRKINRSLSKADSPSRSTTVGL 400  
Db 406 VGNVEYVNGQVVOYIYVTCALAKAVYERFMWYTRINATLETQP-----XQYFIGV 459  
Qy 401 LDYGEVFNHNSFEQFCINVCNEKLOQLFIELTLKSEOEYEAEGIAMEPVOYFNKKII 460  
Db 460 LDIAGFETFDNSFEQFCINFEKLEQOOFNNHMFVLEQEEYKKEGIEWTFTDXGMDLOA 519  
Qy 461 C-DLVEEKFGIISTLDEECRLRGEATDTFLEKLEDTYKPHHFLTHKLADQKTR--KS 517  
Db 520 CIDLI-EKPMGIMSTLEECMFP-KATDPTFKAKLFD-----NHLGKSANFQKPRNIGK 571  
Qy 518 LDRGEFRLHYAGEVYVSTGFLDKNNDLFEKLEKTCSSMNPIMAOCF----- 567  
Db 572 KPEAFSLIHAXIYXVNIIGMLQKNKDRPLXETVAGLVOKSSXLSTLTFAYVACADAPI 631  
Qy 568 DKSELSDKKRP--ETVATQFKMSLLQVLEILSKREPAVIRICIKPNDAKOPGRFDEVILRH 625  
Db 632 EKGKAKKAGSSFOYVSLAHRENKLNKLTMTNLSHPHEVRCTIIPNETKSPGVMDPVLVNH 691  
Qy 626 QYKYLGLMENLVRRAAGAVRRKYPAFLQRYKSLCPETWPMAGRPQD---VAIVLVRHL 682  
Db 692 QLRKGVLEGIRICKRGPNRLIYDFQRYXILNPA--IXEGQFXSRKXAEKLLSSL 749  
Qy 683 GYKPEYKMGRTKIFIRPPKTLFATEDSLLEVROS-LATKIOAMRGFMROKFLRV--K 739  
Db 750 DIDHNYQFGHTKV--FKAGLLGLEMRKXERLSRIITRIDQASRGVLAHRMEYKLLER 807  
Qy 740 RSAICIQSM 748  
Db 808 RDSLVLQW 816

RESULT 3  
US-09-172-422-1  
Sequence 1, Application US/09172422A  
Patent No. 6300485  
GENERAL INFORMATION:

APPLICANT: Adams, Arwen E.  
APPLICANT: Chiu, Choi Ying  
APPLICANT: Duhli, David  
APPLICANT: Gorman, Susan W.  
APPLICANT: Leng, Song  
APPLICANT: Sheffield, Val  
APPLICANT: Welch, Juliet  
TITLE OF INVENTION: MYOSIN IXA AND CYCLIC NUCLEOTIDE GATED  
TITLE OF INVENTION: CHANNEL-15 (CNC-15) POLYNUCLEOTIDES, POLYPEPTIDES,  
FILE REFERENCE: 200130.442  
CURRENT APPLICATION NUMBER: US/09/172.422A  
CURRENT FILING DATE: 1998-10-14  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 2548  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-172-422-1

Query Match 20.5%; Score 1105; DB 4; Length 2548;  
Best Local Similarity 25.2%; Pred. No. 1.4e-95;  
Matches 320; Conservative 205; Mismatches 370; Indels 376; Gaps 32;

Qy 40 SEAAFIENLRFRFRNLNLYTYIGVPLVSNVPRDQIYSRQMERYRGSPFEVPHFLFA 99  
Db 158 NEKTLLENLRDFEKHEKITYTYGSLIYINPEKFLIYTPKVKMIDNHQJLQPREHYTA 217  
Qy 100 VADTVYRALTRERROAAVMISGESAGTEATKRLLOFAETCPAPERGA--VYDRILQ 157  
Db 218 VADVAYHAMLRKKKQCIYISGESSGKTQSTNPLH--NLTALSQKCFASGVEDIILG 274  
Qy 158 SNPVLAEFNMAKTLRNDSSRRGKYMDVOFPDKGAPVGCHILSYLLEKSRVYHONHGERN 217  
Db 275 AGPVLEAFGNMAKTAHNNSSRRGKFIQVNYQETGVLAGYVEKYLLEKSRVLYQEHNEBN 334  
Qy 218 FIVEYQELLEGEETLRRLGLERNPOSYIY-----VKQCAKAVSSI 259  
Db 335 YHVFYLLLAGASEDEERSAHL-KQPEEYHYLNQITRKPLRQSMDDYCYDSEDCFTVEGE 393  
Qy 260 NDKSPMKVYKRALSYIDFTEDEVEDLSTIVASVHLGNHIFAAD--EDSNAQVTTENOLK 317  
Db 394 DLRHDFERQJLAMEVNGFLPKTRRQIFSLSLIHLGNICYKKTTRDSDIDICNPEVLR 453  
Qy 318 YLTRLGVEGTTLRREALTHRKIIAKGEELLSPNLDEQAAYRADALAKAVYSRTFTWYRK 377  
Db 454 IYSELLEVVEEMLFALVATYRKTVTGKELILPYKLAEAVTVNSMAKSLYSALFDWIVFR 513  
Qy 378 INRSILA-SKDASPSRSTTVGLDIYGEVFNHNSFEQFCINVCNEKLOQLFIELTLK 436  
Db 514 INHALNLSKDLEHNT-KTLSIGVLDIFGEDEYENNSFEQFCINFEANETLOHYFNQHIK 571  
Qy 437 SPOEEYEAEGIAMEPVOYFNKKIICDLVEEKFGIISTLDEECRLRGEATDTFLEKLEDT 496  
Db 572 LEOEBYRTBGISWHNIDYIDNTCCINLISKPTGGLHLHDEESNRP-QATNOTLADKFR- 629  
Qy 497 TVKPRPHFLTHKLADQKTR--SLDRGEFRLHYAGEVYVSTGFLDKNND-----LL 547  
Db 630 -----HQHEDNSYIEFPVMEPAIILKHVAGKYVGYKDFREKNTDMMRPDIYAL 679  
Qy 548 FNLKLETCSSM----- 559  
Db 680 LRSSKNAPISGMIGIDPVAVFRMAILRAFFRAMVAFREAGKRNIHRTGHDTPARCALIK 739  
Qy 560 -----NP-----IMAO-----FD-- 568  
Db 740 SMDSEFLQHPVHQSLETLQCKKEKYSITRKNPRTPLSDILOGANALNEKNOHTFDIA 799  
Qy 569 -----KSELSD----- 574  
Db 800 WNGRTGIRQSRLLSSGSLLDKDGIIFANSTYSKLLERAHGILTRNNKFKSPALPKHLEV 859  
Qy 575 -----KKRPETVATQFKMSLLQVLEILSKREPAVIRICIKRN 610  
Db 860 NSLKLTRLTLLODRITKSLHLHKKRKKPSISAOQASLSKMETLGOAEPFYKQIRSN 919  
Qy 611 DAKQGRFEDEVILRHQVYKYLGLMENLVRRAGFAYRRKYPAFLQRYKSLCPETWPMAGR 670  
Db 920 AKKLPLRSDVLYLQKRTGMELEYQIROSYSKYSFODPVSHFVLLPNNIIPSKN 979  
Qy 671 PODGVAVILRHLYGKPEEYKMGRTKIFI-----RFPKTLPA 706  
Db 980 IQD-----FFRKINLMPDNDYQKTVWFLKEQERQHLQDHLQEVLRRIILLRWMPFVLLC 1035  
Qy 707 TEDSLFVRROSL-----ATKIOAMRGFMHROKFL 736  
Db 1036 ROHFLHLRQASVYIQRFWENYLNOKOVRDAVQKDAFVVASAALQASWRAHLEORVY, 1095  
Qy 737 RYKRSALCICQSMWRTGIGRKA-----KKMA-----QITIRLI-----GGLIR 778  
Db 1096 ELRAAAIYIQKRWQYRRKRHAAICIQARWKAAYRESKRYQEQORCKTILLQSTCRQFRAR 1155  
Qy 779 HSPRC-----PENAFILDHVRASFLNLRLQPLRNVLDTSWTPPPALREASELL 828

Db 1156 QRFALKEQRLREKPEV--LVNKGYSLETSQSDPEWEDCSFDRIRKALECKSVI 1213  
QY 829 REICMKMMWVKYCRSTISPEKQOOLQKAVASEIFKQKNDYQSPRLIFISTRLETS 888  
Db 1214 ESNRISSESSVDLKEKSPKQOEKRAQSGVDL---QED-----VLVER----- 1255  
QY 889 PVALQSLGSEPIQYAVPVKCKYKPRRQLLTPSAVYIVEDAKYKORIDYANLTGI 948  
Db 1256 PRSLIEDLHQKKVGR-----KESRRMRELEQALFSLLELKYR-----SLGCI 1298  
QY 949 SVSLSDSLFLVHQREDKQGDVY---LQSDH-----VIFELTKLALSDRVN 995  
Db 1299 SPS-----EDRRWSTELVPEGLQSPGCTPDSSSGSLELLSYESQSKLE 1345  
QY 996 NININGSTTF 1006  
Db 1346 SVISDEGDLQF 1356

RESULT 4  
US-08-938-105-3  
Sequence 3, Application US/08938105  
Patent No. 6353151  
GENERAL INFORMATION:  
APPLICANT: Lelwand, Leslie A.  
APPLICANT: Vikstrom, Karen L.  
TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln St., Suite 3500  
CITY: Denver  
STATE: CO  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/938,105  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Crook, Hannell W.  
REGISTRATION NUMBER: 31,071  
REFERENCE/DOCKET NUMBER: 3595-4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1886 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-938-105-3

Query Match 18.9% Score 1018; DB 4; Length 1886;  
Best Local Similarity 32.18; Pred. NO. 1.8e-87;  
Matches 253; Conservative 136; Mismatches 286; Indels 114; Gaps 17;  
QY 4 BASALSGDCVRVME-----SALTARDRV-----GVQDFVLENTSEAAFTENLR 49  
Db 47 KAKIVSRGCGKVTATETENGKTVTKEDQVMOQNPCKDKIEDAMT-TFLHEPAVLVYLNK 105  
QY 50 RFRFENLIYTYIGVIVSVNRYRDLQIYSQIHMERVGVSEFVPPHILFAVADTVYRALR 109  
Db 106 ERYAAMMITYTSGLFCTVNVYTWLPLVYNAEVAAYRGKKRSEAPRHISISIDNAYQVWL 165  
QY 110 TERDQAVNMISGESGAGCTEATKRLQFYAETGCPAPERG-----GAVDRLLQSNP 160

Db 166 TRENQSTILITGESSAGKTVMTRKVIQYFASIAIGRSGKDDPNANKGILEQIIGANP 225  
QY 161 VLEAFGNKTLKNDNSNFRGKMDVOPDFGAPGVGHILSLYEKRVYVHONGERNFY 220  
Db 226 ALEAFGNKTYVRNNSNFRGKIRIHGATGKLASADIEYILEKSRVIFOLKAEERNYH 285  
QY 221 FYQLLEGGEETLRIGLERPQSYLYLVKGGCAKVSINDKSDMKVMKALSVIDFTED 280  
Db 286 FYQLSNKRPPELLMLLVYNNPYDYAFVSGE--VSAVSIDSEELLALDASADVLGTA 344  
QY 281 EVELDLSTIVASVLHGNHFADE--DSNAOVTTENQKLYTLTLLEGVETTLREALTHRT 339  
Db 345 EKAGVYKLTGAINHYGKMKFKQREGEAEPDTEADKSAVLKGLNSADLLKGLCHPOV 404  
QY 340 IAKGEELSPLELQEAAYADALAKAVYSRTTWTLYRKINSLSAKDAESPMSRTYVLG 399  
Db 405 KVGNEYVTQGSQVQYVYISIGALASVEKMFNMWTRINATLETQPRQ-----YFIG 458  
QY 400 LLDIYGFEPVQHNSEFQFCINYNCKEQLQLEITLTKSEGEYEAEIAMEPVQYFNKI 459  
Db 459 VLDIAGEF-----  
QY 460 ICDLVEKFKGIISLDECLRPGEATDLPLEKLEDTYKPHHFLTKLADQTR--KS 517  
Db 467 ISSPLHLKLMGIMSLTEECHEP--KATDMTFKAKLYD-----NHLGKSNFQKPRNVG 519  
QY 518 LDRCGEFLIHTAGEVYTSVTGFLDKNNDLLEFRLNKETMSSNMPIAAOCPDSELS 574  
Db 520 KQEAHFSLVHAGVYDYNILGMLKKNKDPLETFYGLIQSSSLKLAATLFTSYASADTGD 579  
QY 575 -----KKRP---EYVATQFKMSLDLVLKSKDEPAYTRICIRPDNAKQRFDEVILR 624  
Db 580 SGKGGKGGKSSQYVSAHRENLKMLNTLTTPHVFVRIIPNERKAPVMDNPVLM 639  
QY 625 HOVKYLGLEMLRVRAGAVRRRYEAFLORYKSLCPETPMWMAARPD---GAVLVYRH 681  
Db 640 HOLCNQVLEGIIRCKRGFPNRIYLGDFQRYRIINPAIP--EQQFDSKSGAEKLGS 697  
QY 682 LGYKPEEYKMGRTKIFIRFPKTLFATEDSLVYRROSLATKTIQAMRGFMHOKFLRY--K 739  
Db 698 LDIIDHQYKFGHTKVFYK--AGLLGLEEMRDERLSRIITRIQAOARGLMIRKKNYER 756  
QY 740 RSAICIQSW 748  
Db 757 RDALVYIQW 765

RESULT 5  
US-09-012-515A-12  
Sequence 12, Application US/09012515A  
Patent No. 6127521  
GENERAL INFORMATION:  
APPLICANT: Berlin, Vivian  
APPLICANT: Chiu, Maria Isabel  
APPLICANT: Cottarel, Guillaume  
APPLICANT: Damaguez, Veronique  
TITLE OF INVENTION: IMMUNOSUPPRESSANT TARGET PROTEINS  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLER, HOAG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109-2170  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/012,515A



```

OY 223 Q--LLEGGEETLRRLGLEBNPOSYLVKGCACAVSSINDKSDMKVAKALVIDETED 280
Db 493 QHMLRSGGODALASGPVETGPMKKLH-----VSTINQOKMAAAR----- 534
OY 281 EVEDLSTIASVLIHGINHFAADEDSNAQVTENQKYLTRLGLGVETTTLREALTHRKII 340
Db 535 -----VSKDMLWMLRL----- 547
OY 341 AGEELLSPLNEQAAYARDALAKAVYSRTFTLVKIRINSLASKDAESPMSRTTVLC- 399
Db 548 --SLELL-----KSSSPSLRSCMALAQ 568
OY 400 ----LDIYGFVEFOHNSFEQFCINYNCKEKLQOLF--IELTLKSEGEYEAEGIAWER 451
Db 569 AYNMARDLF-----NAFVSCWSELNEDQODELIRISIELALTSOD-----IA-EV 613
OY 452 VOYFNKKITICDLVEEKFKIISILDE-----ECLRPGEATDITLFEKLEDTVVKP 500
Db 614 TOTLLN--LAEFMEHSDKGPLRLRDNGIVLGERAAACRAVAKA--LHYKELEFOKGP 668
OY 501 HPHFL-----THKLADQKTRKSLDRGEFRLHYAGEVTVSVTGF----- 539
Db 669 TPALLESLSINKKL--QOPEAAAGVLEYAMKHF--GELEIQATWYERKLEHEDALVAAYDK 725
OY 540 -LDKNNDLFRNLKETGSSNMPIMA-----OCFDSKSELDDKRPETVAQOFKMSL 589
Db 726 KMDTKND-----DPELMGRKRCLEALGEMQOLHOCCCEKWTLVN--DETQAKKARMAA 777
OY 590 LQVLIELSKREPATYIRCIKRPNDAKQGRFDEVILIRIQ-----VYUIGLMEELNVR 639
Db 778 AAMGLGOWDSMEETCTMIRPDTHDCAFYRAVALAHQDLFSLAQCCIDKAROLDALDELTA 837
OY 640 RAGFAVYRKYEAFLORYKKSICREPTEMAGRPQDGAVALVLRHLGKRSALC--YKMGRTKI 696
Db 838 MAGESYSRAVGAWS-----C-----HMLSELEVIQYKL----- 867
OY 697 FIRPKTLFATEDSLEVRQSLATKIQAMRGFHWKQFLYKRSALC--IOSMWRGT 752
Db 868 -----VPERREILQIMMERLOCCQIRVEDMOKILMVRSLVSPHEDMTW----- 913
OY 753 LGRRAKAKRMAAQITRRLIRGFLIRHSPRCPENA--FFLDHVRASFLNLRLQOLPRVLD 811
Db 914 -----LKVASLGGSGRLALAKHTVLVLLGVD--PSROLD 946
OY 812 TSMPTPPALREASLELRELCKMMVWKYCRSISP-----EMQOOLQAKAVASEIFK 863
Db 947 HPLPTVHPQVTVYA-----YKMN--KKSARKIDAFQHMONGVOTMQOQOAHAIATEDQO 998
OY 864 CKRDYPOSVPRLFISTRLGTEELSPVLOSLSERP--QYAVRVVUYDKKGYKRPFR 919
Db 999 IKOVLH--KLMAKCFL--KLGEWOLN--LOGINESTIPKVLQYYSATTEHDSMTKAWHA 1052
OY 920 QLLLTSPASV-----IVEDAKVQR-IDVANLTGISVSLSDSLFVLHVOREDKKOKGD 972
Db 1053 WAMNFEVAVLHYKHQONARDEKKKLRHAGANTITNATTAATAAATVTTASTEGSNSSE 1112
OY 973 VVLQSD-----HVIEITLTK-----ALSADRYNNI 997
Db 1113 AESTENSPRPSPLQKKVTEDLSKTLTLMYTVPAVQGFPSRISLSGNL 1160

```

```

RESULT 7
US-09-012-504A-12
; Sequence 12, Application US/09012504A
; Patent No. 6464974
; GENERAL INFORMATION:
; APPLICANT: Berlin, V.
; APPLICANT: Chlu, I.
; APPLICANT: Cottarel, G.
; APPLICANT: Damagnez, V.
; TITLE OF INVENTION: IMMUNOSUPPRESSANT TARGET PROTEINS
; FILE REFERENCE: APBI-P05-036

```

```

; CURRENT APPLICATION NUMBER: US/09/012,504A
; CURRENT FILING DATE: 1998-01-23
; PRIOR APPLICATION NUMBER: 08/360,144
; PRIOR FILING DATE: 1994-12-20
; PRIOR APPLICATION NUMBER: 08/250,795
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 1809
; TYPE: prt
; ORGANISM: Mammalian
US-09-012-504A-12

Query Match
Best Local Similarity 19.18; Score 160; DB 4; Length 1809;
Matches 193; Conservative 121; Mismatches 318; Indels 376; Gaps 46;

OY 137 FYAETCAFERGAVR--DRLQS-----NPVLEAFGNAKTLRN--DNSS-- 177
Db 382 FDAEAPLPKRAKALETVDRLETESLDTFYASRIITHPTVRLDQSPELRSTAMDTSSLY 441
OY 178 -REGKYMDVQDFDKGAPVGGHILSYLLEKSRVYHQN-----GERNFHY 222
Db 442 FOLGKKYQIFIP-----MVKVLVHRIHQRYDVLCRIYVGYTLADEEDPLTY 492
OY 223 Q--LLEGGEETLRRLGLEBNPOSYLVKGCACAVSSINDKSDMKVAKALVIDETED 280
Db 493 QHMLRSGGODALASGPVETGPMKKLH-----VSTINQOKMAAAR----- 534
OY 281 EVEDLSTIASVLIHGINHFAADEDSNAQVTENQKYLTRLGLGVETTTLREALTHRKII 340
Db 535 -----VSKDMLWMLRL----- 547
OY 341 AGEELLSPLNEQAAYARDALAKAVYSRTFTLVKIRINSLASKDAESPMSRTTVLC- 399
Db 548 --SLELL-----KSSSPSLRSCMALAQ 568
OY 400 ----LDIYGFVEFOHNSFEQFCINYNCKEKLQOLF--IELTLKSEGEYEAEGIAWER 451
Db 569 AYNMARDLF-----NAFVSCWSELNEDQODELIRISIELALTSOD-----IA-EV 613
OY 452 VOYFNKKITICDLVEEKFKIISILDE-----ECLRPGEATDITLFEKLEDTVVKP 500
Db 614 TOTLLN--LAEFMEHSDKGPLRLRDNGIVLGERAAACRAVAKA--LHYKELEFOKGP 668
OY 501 HPHFL-----THKLADQKTRKSLDRGEFRLHYAGEVTVSVTGF----- 539
Db 669 TPALLESLSINKKL--QOPEAAAGVLEYAMKHF--GELEIQATWYERKLEHEDALVAAYDK 725
OY 540 -LDKNNDLFRNLKETGSSNMPIMA-----OCFDSKSELDDKRPETVAQOFKMSL 589
Db 726 KMDTKND-----DPELMGRKRCLEALGEMQOLHOCCCEKWTLVN--DETQAKKARMAA 777
OY 590 LQVLIELSKREPATYIRCIKRPNDAKQGRFDEVILIRIQ-----VYUIGLMEELNVR 639
Db 778 AAMGLGOWDSMEETCTMIRPDTHDCAFYRAVALAHQDLFSLAQCCIDKAROLDALDELTA 837
OY 640 RAGFAVYRKYEAFLORYKKSICREPTEMAGRPQDGAVALVLRHLGKRSALC--YKMGRTKI 696
Db 838 MAGESYSRAVGAWS-----C-----HMLSELEVIQYKL----- 867
OY 697 FIRPKTLFATEDSLEVRQSLATKIQAMRGFHWKQFLYKRSALC--IOSMWRGT 752
Db 868 -----VPERREILQIMMERLOCCQIRVEDMOKILMVRSLVSPHEDMTW----- 913
OY 753 LGRRAKAKRMAAQITRRLIRGFLIRHSPRCPENA--FFLDHVRASFLNLRLQOLPRVLD 811
Db 914 -----LKVASLGGSGRLALAKHTVLVLLGVD--PSROLD 946
OY 812 TSMPTPPALREASLELRELCKMMVWKYCRSISP-----EMQOOLQAKAVASEIFK 863
Db 947 HPLPTVHPQVTVYA-----YKMN--KKSARKIDAFQHMONGVOTMQOQOAHAIATEDQO 998

```

QY 864 GKDNVPOSVPRLFTSTRGTEISPRVLOSISSEPI---OYAVPVYKDRKGYKPRPR 919  
 Db 999 HKQELH-KLMARCLF--KLGEWQLN---LOGINESTIPKVLQYSAATHEHDSRWKAMHA 1052  
 QY 920 QLLTPSAVY-----IYEDAKVKOR-IDYANLTGISVSLSDSLFVLHVOREDNKOKGD 972  
 Db 1053 WAWMFEAVLVHKHONQARDEKKLRHAGANITNATTAATTAATTTASTEGSSESE 1112  
 QY 973 VVLQSD-----HVETLTKT-----ALSADRVNNTI 997  
 Db 1113 AESTENSPPLQKKVTEDLSKTLMTVPVAVGCFRFSISLRGNL 1160  
 RESULT 8  
 US-08-471-112A-3  
 ; Sequence 3, Application US/08471112A  
 ; Patent No. 6313264  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Molnar-Kimber, Katherine L.  
 ; APPLICANT: Failli, Amedeo F.  
 ; APPLICANT: Caggiano, Thomas J.  
 ; APPLICANT: Nakanishi, Koji  
 ; APPLICANT: Chen, Yangu  
 ; TITLE OF INVENTION: EFFECTOR PROTEINS OF RAPAMYCIN  
 ; NUMBER OF SEQUENCES: 23  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
 ; ADDRESS: Duner, L.L.P.  
 ; STREET: 1300 I Street, N.W.  
 ; CITY: Washington  
 ; STATE: DC  
 ; COUNTRY: USA  
 ; ZIP: 20005-3315  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/471,112A  
 ; FILING DATE: 06-JUN-1995  
 ; CLASSIFICATION: 536  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/384,524  
 ; FILING DATE: 13-FEB-1995  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/312,023  
 ; FILING DATE: 26-SEP-1995  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/207,975  
 ; FILING DATE: 08-MAR-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Slekmann, Michael T.  
 ; REGISTRATION NUMBER: 36,276  
 ; REFERENCE/DOCKET NUMBER: 01142, 0058-00000  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 202-408-4000  
 ; TELEFAX: 202-408-4400  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2549 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; US-08-471-112A-3  
 Query Match 3.0%; Score 160; DB 4; Length 2549;  
 Best Local Similarity 19.1%; Pred No. 3.7e-05;  
 Matches 193; Conservative 121; Mismatches 318; Indels 376; Gaps 46;  
 QY 137 FYAETCPAPERGAVER--DRLIQS-----NPVLEAFGNKTLRN--DNSS--- 177

Db 1122 FDAPAPLPFSRKAALETYDRLTESLDFDYASRIIHPHYKRLDQSPBLRSTAMPDLSSIV 1181  
 QY 178 -REGKYMDOGFQFKGAPVGNHILSTYLEKSRVYVHONH-----GERNPHYV 222  
 Db 1182 FOLGKKYQIIFP-----WNVKVLVNRHINRHOYDYLGRIRYKGYTLADEEDPLIY 1232  
 QY 223 Q--LLEGGEEETLRRLGLERNPOSYLTVKGGCAVSSINDKVMKALSVIDFED 280  
 Db 1233 QHRMLRSGGGDALAGPVEYTGPMKKLH-----VSTINLQKANGAAR----- 1274  
 QY 281 EYEDILSTIVASVLHGNHFAADEDSMAOYTTENOLKYITRLGVBGTTLRREALTHRKII 340  
 Db 1275 -----VSKDMLLEWLR----- 1287  
 QY 341 AKGEELSPNLNQAAAYARDALAKAVYSTFTWLRKINRSLASDAPSPSRSTTVLG- 399  
 Db 1288 --SLELL-----KDSSPSLRSCWALAQ 1308  
 QY 400 -----LDIYGFEVFOHNSFEQFCINCEKIQOLF---IELTKSEGEYEBAEGIAMP 451  
 Db 1309 AYNPARDLF-----NAAFVCSNELNEDQODELIRSELALISQD-----IA-EV 1353  
 QY 452 VOYFNKKIICDLVEEKFGIISILDE-----ECLRGEATDLTEKLEDPYKP 500  
 Db 1354 TQTLN--LAFFMESHDKPLPLRDNGIIVLGERAKRACRAYAKA---LHYKELEFQKP 1408  
 QY 501 HHFL-----THKLADQKTRSLDRGEFRLIHYAGEVYVSTGF----- 539  
 Db 1409 TPALLESLSISINNKI--QPEAAAGVLEYAMKH--GELEIQATVEKLEHWEDELVALYDK 1465  
 QY 540 -LDKNNDLFRMLKETCWSMNPIMA-----QCFDKSEISDKRPEYATQFKMSL 589  
 Db 1466 KMDTKND-----DPELMLOMRGLEALGEMQLOHQQCEKMTLVN---DETQAKKARMA 1517  
 QY 590 IQLVELLSKEPARYIRCIKRPNDAKOGPRDEVILRHQ-----VKYGLMENLVR 639  
 Db 1518 AAAMGIGOWDSMEEYTCMIPRDTHGAFYRAVALAHQDLFSLAQCIDARLDLDELTA 1577  
 QY 640 RAGFAYRREYEAFIQRYKSLCPETPMWAGRODGAVALVRLHGYKPER--YKMRRTI 696  
 Db 1578 MAGESYRAVGAWS-----C-----HMLSELEVIOKTL----- 1607  
 QY 697 FIRPKTLFATEDSLVFRSOSLATKIOAAMRGFMHROKFLRYKRSATC-----IOSWMBGT 752  
 Db 1608 -----VPERREITIQIMMERLOGGORLYEBMQKILMRSLSVSHEDMRTN----- 1653  
 QY 753 LGRRAAKRMAAQITRRLIRGLTILHSPRCPENA--FLIDHYRASFLNLRQLPNNVD 811  
 Db 1654 -----LKYASLCGSGRLALAHKTLVLLGVD--PSRLD 1686  
 QY 812 TSWPTPPRALREASSELLRELCKKNYWKYCRSISP-----EMKQOIQOAVASEIRK 863  
 Db 1687 HPLPYVHPQVYA-----YKMKM-WKSARKIDAFQHNQHFVQTMQOQOAHAIATEDQ 1738  
 QY 864 GKDNVPOSVPRLFTSTRGTEISPRVLOSISSEPI---OYAVPVYKDRKGYKPRPR 919  
 Db 1739 HKQELH-KLMARCLF--KLGEWQLN---LOGINESTIPKVLQYSAATHEHDSRWKAMHA 1792  
 QY 920 QLLTPSAVY-----IYEDAKVKOR-IDYANLTGISVSLSDSLFVLHVOREDNKOKGD 972  
 Db 1793 WAWMFEAVLVHKHONQARDEKKLRHAGANITNATTAATTAATTTASTEGSSESE 1852  
 QY 973 VVLQSD-----HVETLTKT-----ALSADRVNNTI 997  
 Db 1853 AESTENSPPLQKKVTEDLSKTLMTVPVAVGCFRFSISLRGNL 1900  
 RESULT 9  
 PCT-US95-06722-12  
 ; Sequence 12, Application PC/TUS9506722  
 ; GENERAL INFORMATION:  
 ; APPLICANT:

```

: TITLE OF INVENTION: Immunosuppressant Target Proteins
: NUMBER OF SEQUENCES: 25
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII (text)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/06722
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/250,795
: FILING DATE: 27-MAY-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/250,795
: FILING DATE: 20-DEC-1994
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2549 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US95-06722-12

```

```

Query Match      3.0% Score 160; DB 5; Length 2549;
Best Local Similarity 19.1%; Pred. No. 3.7e-05;
Matches 193; Conservative 121; Mismatches 318; Indels 376; Gaps 46;

```

```

QY 137 FYATCPAPRGAVR--DRLOS-----NPLAEFGNAKTRN---DNSS--- 177
DB 1122 FDRPAPLPSSKAALETWDLRESLDFDYASRIHPVRLDPSPELRSAMDJLSLV 1181
QY 178 -REGKMDVQDFPKGAPVGGHLLSYLLEKSNVHONH-----GERNFHVEY 222
DB 1182 FOLGKKYQIFIP-----MVKVIVRHINRORTDVLICRTVKYTLADEEDPLIY 1232
QY 223 Q--LLEGESEETLRRLGLERPQSYLYLVKGQCAKSSINDKSDMKYMRKALSVIFTE 280
DB 1233 QHRMLRSGCGALASGPVEGTGPMKKLH-----VSTINLQKAMGAAR----- 1274
QY 281 EVELDLSTVASVLHGINHFAADEDSNAQYTTENOLKYLRLLGVCITLREALTHRKII 340
DB 1275 -----VSKDMLLEMLRL----- 1287
QY 341 AKGEELSPNLLEQAAVARALAKAVYSRTFTLVKRNKRSLSAKDAESPSWRSTTVLG- 399
DB 1288 --SLELL-----KDSSEPSLSKSMALAQ 1308
QY 400 -----LLDIYGFVFOHNSFEQFCINVCNEKLOQLF---IELTLKSEDEEVEAGIAWEP 451
DB 1309 AYNPMARDLF-----NAFVSCWSELNEDODDELINSIELALTSOD-----IA-EV 1353
QY 452 VQYFNKKITCDLVEEKFKITISLDE-----ECLEPGEATDLTFLEKLEIDVYKP 500
DB 1354 TOTLLN--LAETHEHSDKGPLRLRDNGITVLGERAKRCRAYAKA--LHKLELEFGKP 1408
QY 501 HHFFL-----THKLADQKTRKSLDRGFEFLHYAGEVTVSYGF----- 539
DB 1409 TPALESLSISNNKL--QOPEAAQVLEYAKKH--GELETOATWYERLHMEEDALVAYDK 1465
QY 540 -LDKNNDLFLRLKFTKSSNMPIMA-----QCFDSELSDDKKRPETVATQFKMSI 589
DB 1466 KMDTNKD-----DEPLMGRNRRCLEALGEGQLHQOCCCKWTLVN--DETQAKKARMAA 1517
QY 590 LQVLEILSKPEAVYIRCIKPNDAKQPRFDEVLIRHQ-----VYTLGIMENLVRV 639
DB 1518 AAANGLCQWDSMEYETCMIPDRFHDAFYRAYLALNODLFSLAQOCIDAROLDLAEIETA 1577
QY 640 RAGFAVRYKYEAFIQRYKSLCPEWPMWAGRPDGVAVLVRLHYKPEE---YKNGRTKI 696
DB 1578 MAGSFSRAVGAWS-----C-----HMLSELEVIQYKL----- 1607

```

```

QY 697 FIRFPKTLFATEDSLVEVROSALTKIQAAHGFHWROKFLVRSALC-----IQSWRGT 752
DB 1608 -----VPERREITRQIMWERLOGCORIYEDWOKILMVASLYVSPHEMRITM----- 1653
QY 753 LGRKAARKKMAAOTIRRLNGFLRLRSPPRCPEHA--FELDHYVASLMLRLRQLRNVLD 811
DB 1654 -----LKVASLCSGSGRLALAHKTLVLLGVD---PSRQLD 1686
QY 812 TSWPTPPALREASELLRELCMKNMVWKYCRISIP-----EKKOOLQKAVASEIFK 863
DB 1687 HPLPTVHPQTYA-----YKNNM--WKSARKIDAFQHMQHVFQTMQOAOAHATATEDQ 1738
QY 864 GKNDNPQSVPRFLFISTRIGTEISPRVQSLGSEPI-----QYAVPVYKRYKGYKPPR 919
DB 1739 HKQELH-KIMARCFL--KLGEWQLN--LQGINESTIRKVLQIYSATHEHRSWYKAMHA 1792
QY 920 QLLLTSPSAVY-----IYDAKVKOR-IDYANLTGISVSSLSLSLFLVHQREDNKQGD 972
DB 1793 WAVNFEAVLVHYKQNOARDEKKLRLHASGANINATTAATTAATTTASTEGSNESE 1852
QY 973 VVLQSD-----HVLETLTKT-----ALSADRVNMI 997
DB 1853 AESTENSPSPSPLOKKVTEDELSTKLMLMTYVPAVGFFRSISLSRGNNL 1900

```

```

RESULT 10
US-08-905-223-395
: Sequence 395, Application US/08905223
: Patent No. 6222029
: GENERAL INFORMATION:
: APPLICANT: Edwards, Jean-Baptiste D.
: APPLICANT: Duclert, Aymeric
: APPLICANT: Lacroix, Bruno
: TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS
: NUMBER OF SEQUENCES: 503
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobe, Martens, Olson & Bear
: STREET: 501 West Broadway
: CITY: San Diego
: STATE: California
: COUNTRY: USA
: ZIP: 92101-3505
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy Disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Word
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/905,223
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Israelsen, Ned A.
: REGISTRATION NUMBER: 29,655
: REFERENCE/DOCKET NUMBER:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 235-8550
: TELEFAX: (619) 235-0176
: INFORMATION FOR SEQ ID NO: 395:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 103 amino acids
: TYPE: AMINO ACID
: TOPOLOGY: LINEAR
: MOLECULE TYPE: PROTEIN
: ORIGINAL SOURCE:
: ORGANISM: Homo Sapiens
: TISSUE TYPE: Brain
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: -72..-1
: IDENTIFICATION METHOD: Von Heijne matrix
: OTHER INFORMATION: score 3.6
: OTHER INFORMATION: seq CTSLLQLYDASNS/EW

```









QY 130 ATKRLQFYAETCPA-PERGAVRDRLLOS--NPVLE-AFGNAKTLNDNSSRGKIMDV 185  
Db 51 ---LILQNLNMLCOAOPRSASDVEEROKSFPHIDKMAIADAQSAIEKRRK----- 100  
QY 186 QFDFKCAPVGCILSYLLEKSRVHONHGERNHVFQ-----LIEGGEETLRLRLG 238  
Db 101 -----NPLS--LPAEKIHHLLREYLGKIDHVSVYVAVLEYIASDILKLVG- 146  
QY 239 ERNPOSYL-----YLVKQCAKYSINDKSDMKVMRKALSVIDFTEDEVEDLSTIVASVL 293  
Db 147 -----NVVRNRIHYEITKQDIKAMCADK-----VLMDFHQDVEDI----- 183  
QY 294 HLCNHPADED-----SNAQVTEN 314  
Db 184 ---NILSLTDEEPSTSGEOTYYDLVKAFAEIRQYIRELNLITKVFREPFVNSKLFSSN 240  
QY 315 QLKYL-----TRLLG-VEGTT--LREALTRKRIKAGEELLSPINLE-QAAVYAR 359  
Db 241 DVENITSRIVDIHELKSVKLGHIEDVEMTDEGSPHPLVGCFEDLAEEIAPDPYESYAR 300  
QY 360 DAL-----AKAVYSRT-----FTWLVRKI 378  
Db 301 DILRPFHGHELSQLSKPGALYLQSIGEGFRKAVOYVLPRLILAPVYHCHLHFELLKQL 360  
QY 379 NSLASKDAESPWSRSTYVLGLDIY-GFEVFOHNSF-----EQFCINYCNE-KLQOLF 430  
Db 361 EEK--SEDEDEKCMQALITALLNVOGMEKISKSIAKRRLESACREYSQOMKGKOLA 418  
QY 431 IELTLSEQEEYAEIGIAMEPVOYFNKKIICD---LYEKKFKGISIID--ECCLR--- 481  
Db 419 IK--KMEIQAKNIDWEGKIDGOCCEFIEMBTLTRVGAKHERHIFLFDGLMICCKSNH 475  
QY 482 -----PGEATP-----DLTLEKLEEDTVKPHPLTHKLADQTKRKSLDGGEFR-----L 525  
Db 476 GQRLPGASSAEYRLKEKPFMRKVQINDKO-----DTSEYKHAFEIT 517  
QY 526 LHVAGEVYTSVYGFLDKN--DILFRNLKETWCSSNPIMAOCEDKSELSDKKRPE 579  
Db 518 LKQNSVILFSKSAEKKNNMMAALISLOYRSTLERM-----LDVTVLOEEKE-- 564  
QY 580 TVATQKMSLLOLVEILRSKPAVIRICINPKNDAKOPGRPE-----VLIRNOVKYGL 632  
Db 565 -----EOMRLPSAEYVRAEP-----DSEENILFEENVOPKAGIPIIKAGTVLKL 609  
QY 633 MENLRYRAGFAVRRKYEAFLOQYKSLCPETWPMAGRPQDGVAVLVRLHLY-KPREYKM 691  
Db 610 IERLYTHM--YADPNFVRFTLTTRSC-----RQBELSLIEREIREPEPTEA 658  
QY 692 GRTKIFIRPKTLFATED-----SLEVR-----OSLATKIQAMRGFHW-RQKFLRYK 739  
Db 659 DR-----IAIENGQPLSALCKRFKREYIQVQLRVLVCR--HWVEHHEFYDFE 705  
QY 740 RSAICTOSM--WRGTLGRKKAARKKMAQTIIRLIR-----GFIILHSPRCPE 785  
Db 706 RQADLLQRMEEFEGTV--RGAAMKMW-VESTYKTIQKRIARDNGPGHNITFOSSP--PT 760  
QY 786 NAFELD--HVRASFLLNLR--ROL-----PRNVLDTSWP-----TPP 819  
Db 761 VEMHISRPGHIEFTFDLLTLPLEIARQLTLESDDLRAVQPSLVGSVMTKEDKEINSPN 820  
QY 820 ALR-----EASLRELCKMKNWKKYCRST-- 844  
Db 821 LKMIHHTNLTLMPFEKCIJETENLEERVAVVSRIEILQVFOELNNFGLLEVVSAMNS 880  
QY 845 SPEWK-----OOL--OQKAVASEIFKGGKDNV-----POSVP--RLFISTRLGT 884  
Db 881 SPVYRLDHTPEQIPSRQKILIEBAHELSEDHKKYLAKLRSINPCVPEFGIYLTNLIKLT 940  
QY 885 EELSPRVLOSLEPIQYA---VPVVKYDRKGKPRPRQLLTPSAVVIVEDAKVKORI 940  
Db 941 EGNPPEVLNRHKGELINFKRRRVAIEITGEIOQYQNOYCLRVEPD-----IKREF 991  
QY 941 DYANLTGISV--SSLSDSLFVLHVQREDNKKOK 970

Db 992 ENLNPNGNSMEKEFTDYLFNKSLEIEPRHKK 1022  
Search completed: December 9, 2002, 16:48:40  
Job time : 37 secs



QY 28 GVDGFDLLNMFNSEAFLEINLRARRREULLITYYGVLVSNPYRDLQIYSRQMERYRG 87  
 Best Local Similarity 38.1%; Pred. No. 14e-99;  
 Matches 265; Conservative 154; Mismatches 248; Indels 28; Gaps 12  
 Db 91 GVEDMSEL-STYNEAPVPHNLKVRNMOLITYYGGLVLVAVNPFKRPIRYTQEMVDFKG 149  
 QY 88 VSFYEVPPLFVADTVYRALTERRDVAVIMSGESAGATEATRRLLQYATCAPER 147  
 Db 150 RRRNEAPAFHFIASDVAVRSMLDKROQNSLITTESAGTENTKKYIQLVASVAGRMA 209  
 QY 148 GGA--VRBRLQSNVLYEAFGNATLKRNDSSRFKGYMDVDFPKAPVGNHTLYLEK 205  
 Db 210 NSSGVLEQOILLQANLLEAFGNATTTNNSSSRGKFEIEIOPNSAGFISGASIQSYLEK 269  
 QY 206 SKVYHQNIGERNFHFVQYLEGGEETLRRLGLEERNOSYLYVKGQCAKVSINDKSM 265  
 Db 270 SKVYFQSETERNHLYFQLLAGATAEKKMLHL-AGRPESNTIMQSGCVDIGVDSDEF 328  
 QY 266 KVMRALSLVDFTDDEVEDLSTVASYVLHGNIFAADSDSNAQVTTENOLUYTLRLGV 325  
 Db 329 KITRAMDLYEFSOEQMSIFKLIAGLHGNIKFEKGAGEGAVLKDKTALMASTVGV 388  
 QY 326 EGTTLREALTRKLIKAGEELLSPNLDEQAARQALAAVYSRFTWLYKRNINSLASK 385  
 Db 389 NPVLEKALMEPRILAGRDLYAQHLVAKSSSRDLALVGRLEFMYLVKINNVLCOE 448  
 QY 386 DAESPSRSTTVGLDIYGFYFOHNSPEQFCINCEKLOOLETELTLSQEEVEAE 445  
 Db 449 -----RKAFYGVLDISGELEFKNSFEOLCINTNEKLDQFPNNHMFYLEDQEVLYKE 501  
 QY 446 GIAPENVQY-FNNKIIQDDVE-EKFYGISILDECELRGEATDLYTFLEKLEDTYKPH 503  
 Db 502 KINMTFIDELDSQATDILIDGRQPGILLALDEQSVF-NATDNTLTITKL-----HSH 554  
 QY 504 FLTHKLADOKTRSLDRGEFRLIHTAGEVTSYVTFEDKNDNDFLRNLKETMCSNMPIM 563  
 Db 555 FSKRNKAYEPRRS--KTEFGVTHAGQWMEIOLMKENNDPQOODLELCFKSDSNV 612  
 QY 564 AQCFDKSELSDKKRP-----TVATQFKMSLLQVLEILSKREPATYRIGCPMDAKOPRFD 619  
 Db 613 TKLENDPNIASRAKKCANETVAQYKQDLASLMAETLETPHVRKITLRNNKOLPAKLE 672  
 QY 620 EYLIRHQVYVYLGMELRYRAGFYRRKRYAEFLQRTKSLCPEFWMAAGRPODGVAVLV 679  
 Db 673 DKVYLDQLKNGVLECIKIRTKRGEPNRLITVADFVKRYLLAPNV-PRDAEDSOKATDAVL 731  
 QY 680 RHLGKPEEYKMGKRTKIFIRPKTLFATEDSLEVR 714  
 Db 732 KHLNDPEQYRGITKIFFRAGQ-LARIEARQR 765  
 RESULT 3  
 US-09-925-300-1583  
 ; Sequence 1583, Application US/09925300  
 ; Patent No. US2002015181A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Craig Rosen,  
 ; APPLICANT: Steve Ruben  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
 ; FILE REFERENCE: P101  
 ; CURRENT APPLICATION NUMBER: US/09/925, 300  
 ; PRIOR FILING DATE: 2001-08-10  
 ; PRIOR APPLICATION NUMBER: PCT/US00/05988  
 ; PRIOR FILING DATE: 2000-03-08  
 ; PRIOR FILING DATE: 1999-03-12  
 ; NUMBER OF SEQ ID NOS: 1890  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 1583  
 ; LENGTH: 569  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens

```
FEATURE:
NAME/KEY: SITE
LOCATION: (2)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (34)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (188)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (291)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (345)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (346)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (552)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (553)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (554)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1583
```

Query Match 22.0%; Score 1186; DB 10; Length 569;  
Best Local Similarity 47.1%; Pred. No. 1.6e-85;  
Matches 243; Conservative 86; Mismatches 167; Indels 20; Gaps 5;

```
Qy 285 LLSIVASVHLGNTHFAEDDSN---AQVTTENOLKLTILGVEGTTLRALTHRKTI 340
Db VLAIVAAVXKLGKIGEFKPSRNVNGDESKIKDKMELKEICELTGIDQSVLEAFRTVE 87
Qy 341 AKGEELSPNLNDOAAVARDALAKAVSRFTFWLVKINRSLASDASPSWRSTTVGL 400
Db 88 AKQEVSTTLNVAQVYARDALAKVLSRLSMLVNRINESTKQTKVRK-----VMGV 142
Qy 401 LDIVGEVFOHNSFEQFCINCNKILQDLFTELKSEQEEAEAGIAMPEVQYFNKTI 460
Db 143 LDIVGEFELFEDNSFEQFIINCNKILQDLFTELKSEQEEFIREXIMETHIDYNNATI 202
Qy 461 CDLYEKEKGIISILDECLRPGEATDLFLEKEDTVKPRPHFLTH--KLADQTKRSL 518
Db 203 CDLIENNNTGLAMDECLRPGEATDLFLEKLNQVCATIQHFESRMSKCSRFINDTSL 262
Qy 519 DRGEFRLHYAGEVTVYTGFLDKNNDLFRNLKETMCSMNPIMAGCFDSELS--DKK 576
Db 263 PHSCRIQHYAGKVLQYQEGEVDKNDLXYRDLQAMMKASIALKSLFPEENPAKTIK 322
Qy 577 RPEVATQFKMSLQDLVEILRSKEPAYIRCIKPNDAKQGRFEVILRHQVYLLMENL 636
Db 323 RPEVATQFKMSLQDLVEILRSKEPAYIRCIKPNDAKQGRFEVILRHQVYLLMENL 636
Qy 637 RVRNAGVAFRRKYEAFLORYKSLCPETPRMAAGRPQGVAVLVRLGKPREYKKGRTKI 696
Db 383 RVRNAGVAFRRKYEAFLORYKSLCPETPRMAAGRPQGVAVLVRLGKPREYKKGRTKI 696
Qy 697 FLPKPKTLEATFEDSLFVROSALATKIOAMRGFMHRKOKFLRVKRSALTIQSMWRGTLLRR 756
Db 443 FLPKPKTLEATFEDSLFVROSALATKIOAMRGFMHRKOKFLRVKRSALTIQSMWRGTLLRR 756
Qy 757 KAAKRMQAQITRLIRLNGF-----ILRHSRCPD 785
Db 503 RYQQTSSALVQSYIRGMKAKIRLELKHOKRCKE 538
```

RESULT 4  
US-09-851-682A-1

```
Sequence 1, Application US/09851682A
Patent No. US20020091248A1
GENERAL INFORMATION:
APPLICANT: Adams, Arwen E.
APPLICANT: Chiu, Choi Yung
APPLICANT: Duhl, David
APPLICANT: Gorman, Susan W.
APPLICANT: Leng, Song
APPLICANT: Sheffield, Val
APPLICANT: Welch, Juliet
TITLE OF INVENTION: MYOSTIN IXA AND CYCLIC NUCLEOTIDE GATED
TITLE OF INVENTION: CHANNEL-15 (CNGC-15) POLYNUCLEOTIDES, POLYPEPTIDES,
FILE REFERENCE: 200130.442
CURRENT APPLICATION NUMBER: US/09/851,682A
PRIOR FILING DATE: 1998-10-14
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 2548
TYPE: PRT
ORGANISM: Homo sapien
US-09-851-682A-1
```

Query Match 20.5%; Score 1105; DB 10; Length 2548;  
Best Local Similarity 25.2%; Pred. No. 3.9e-78;  
Matches 320; Conservative 205; Mismatches 370; Indels 376; Gaps 32;

```
Qy 40 SEAFIENLRFRRENLIYTYIGPVLSVNPYRDQIYSRQHMERVGVSEFVPRHFLA 99
Db 158 NEKTLLENLRDRFKHEKITYYVGSILYINPRKPLIYNPKVKNYDHNQJGKRRPHIYA 217
Qy 100 VADIVYRALRTERRDQAVMISGSGAGCTEATKRLDQFYAETCPAPERGA--VYDRLLQ 157
Db 218 VADIVYHMLQKKNOCIVISGSGAGTQSTNPLIH--HLLTASQKGFASGVQJILG 274
Qy 158 SNPVLEAFGNKTLNDNSRFGKXMDQOFDEKAPGVGHILSYLLEKSRVYVHNGERN 217
Db 275 AGPVLEAFGNATTAHNNNSRFGKTIQVNYOETGVLAGVYKYLLEKSRVYVHNGERN 334
Qy 218 FHFVYOLLEGGEETLRIRGLERNPOSYLYL-----VKQCAKVSIS 259
Db 335 YHVEFYLLAGASEDRSAPFHL-KQPEYHYLNQITKKPLRQSDWDYCYDSEPDCTVBE 393
Qy 260 NDKSPWKYMRKALSYIDTFEVEDLSIVASVHLGNTHFAAD--EDSNAQVTTENOLK 317
Db 394 DLRHDFERLQIAMEVGVLPKTRROIFSLISAHLHGNICYKKTYPKRDSDIDICNPEVLP 453
Qy 318 YLTRLGVGGTTLRRALTRKTIKAGEELSPNLNDOAAVARDALAKAVSRFTFWLVK 377
Db 454 IYSELLEVKKEMLFELALYTRKTVYGEKILIPYKLAEAAYVANSMAKSLYSALFPMIVK 513
Qy 378 INRSIA-SKDAESPMSRSTTVYGLDIYGEVFOHNSFEQFCINCNKILQDLFTELTLK 436
Db 514 INHALNLSKDELEHNT-KTISIGVLDIGFEDYENNSFEQFCINCNKILQDLFTELTLK 436
Qy 437 SEQEEYEAEGIAMPEVQYFNKTIQDLVEKEKGIISILDECLRPGEATDLFLEKLED 496
Db 572 LEQEEYRTEGISMHNIDYIDNTCCINLSKPTGLHLHLEDESPP-QATNQTLDKFK- 629
Qy 497 TYKPRPHFLTKHLADQKRR--SLDRGEFRLHYAGEVTVYTGFLDKNND-----LL 547
Db 630 -----HOHEDNSYIERPAVMEPAFLIKHYAGKVKYGVKDERKNTDHRPDIVAL 679
Qy 548 FNLKETMCSM----- 559
Db 680 LRSSKNAPISGIGIDIPAVFRWALIRAFRRAMVAFRRAGKRNIRKTHGHDGTACALIK 739
Qy 560 -----NP-----IMAC-----FD-- 568
Db 740 SMDSFSLQHPVHQRSLELQRCCKEKKYSITRKNNRPTSLDQGNALNEKNQHDTFDIA 799
```

```

RESULT 5
US-09-925-302-686
; Sequence 686, Application US/09925302
; Patent No. US2002004941A1
GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 686
; LENGTH: 245
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (145)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-302-686
Query Match          10.5%; Score 568; DB 10; Length 245.

```

Best Local Similarity 52.18; Pred. No. 2.2e-37;  
Matches 112; Conservative 33; Mismatches 66; Indels 4; Gaps 2.

```

RESULT 6
US-09-764-853-551
; Sequence 551, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ206
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 551
; LENGTH: 570
; TYPE: PR1
; ORGANISM: Homo sapiens
US-09-764-853-551

```

Query Match	10.5%	Score 566;	DB 10;	Length 570;
Best Local Similarity	27.6%;	Pred. No. 1.2e-36;		
Matches 167; Conservative	91;	Mismatches 198;	Indels 150;	Gaps 20;

```

0Y 359 RDALAKAVYSTFTWLVKRTINRSJAKDAESESWMSTRTVGLDLYTGEVQHNHSFQFC 418
   | | | | | : | | | | | : | | : | : | | | | | | | | |
Db 14 RDCAKLAVILRDLMDVSVINSSIC--ADTDSW--TTFIGLDDVGGESPNDNSLOLC 68
   | | | | | : | | | | | : | | : | : | | | | | | | | |
0Y 419 INYCNKILQDLFIELLKSQOEYKAEKGIAMPEVOYFNKKITCDLVEKFKGIISIDEE 478
   | | | | | : | | | | | : | | : | : | | | | | | | | |
Db 69 INVANERLQOHFAVANHLRQOEYAVEGLEWSEFIYNQONQCLDIIESPISICSLNEE 128
   | | | | | : | | | | | : | | : | : | | | | | | | | |
0Y 479 CL--RGEATDLTEKLEKDJVQPRHNFTHLADOKRKSLDGERLLIHYAGEVTVSY 536
   | | | | | : | | | | | : | | : | : | | | | | | | | |
Db 129 CRLNRPSSAAOQ--TRIFALAGSPCLGNHKLSEBS-----FIVHYAGPRVRYHT 178
   | | | | | : | | | | | : | | : | : | | | | | | | | |
0Y 537 TGFDFKNNLLERLNKETMCSMNPIMAOCSF--DKSELSDKRRE-----TVAYQFM 587
   | | | | | : | | | | | : | | : | : | | | | | | | | |
Db 179 AGVLEKNKPRIPPELTRILLOOQODPLGLGEPNPKETQDEPPGOSRAVLTIVSKFEA 238
   | | | | | : | | | | | : | | : | : | | | | | | | | |
0Y 588 SLLOLVEILRSKEPAYIRCIKPNNAKQGRGRDEVLIHQVYUIGLMENTLRVRRAGFAYKR 647
   | | | | | : | | | | | : | | : | : | | | | | | | | |
Db 239 SLEQQLLOVHSTTPYITICIPNSOGAQOTFLOEYVLSQLEACGLVETIHISAGPIRYV 298
   | | | | | : | | | | | : | | : | : | | | | | | | | |
0Y 648 KEALDQRYKSL-----C-----PET-----WPMWAGRPQD----- 673
   | | | | | : | | | | | : | | : | : | | | | | | | | |
Db 299 SHRNVEERYKLLRLRHPTSGSDPSYIPAKGLREKQCHSEATLEBLDILHTPRVLTQ 358
   | | | | | : | | | | | : | | : | : | | | | | | | | |
0Y 674 GVAVLYRHLGKRPPEYIMKGTKTFIREPKTLFATEDSLRVSQSLATKIQAAMRQFHNRO 733
   | | | | | : | | | | | : | | : | : | | | | | | | | |
Db 359 AALITGDSAEAMPAMHCGRTKYVM-----TDSMELLECGARVLE----- 400
   | | | | | : | | | | | : | | : | : | | | | | | | | |
0Y 734 KFLVRKSAICISQSMWGVTGCRKKAARKKV--AAQITRLIRGFIIRHSPRCENAFELDH 792
   | | | | | : | | | | | : | | : | : | | | | | | | | |

```

Db 401 -----QCARCQCGWR--RRHRREQEROMRAVMIQAIRNSWLT-----KH 440  
Qy 793 VRASELNLRLPNNVLDTSWTPPPALREASELLRELCMKNNWVKYCRSISPENKQDL 852  
Db 441 IQ-----RLHAATVJIKRAMQ-----KKRI 460  
Qy 853 QOKAVASELFKKKNYQOSVPLRFLSTRLGTETSPRYLOS-----LGSEPTIYAVPV 906  
Db 461 RMACIAAAELDCEVKEHESQAP-----CSLSTSPLOTRLLEAIIRFWPLGLVLTAMGV 515  
Qy 907 VKYDRK 912  
Db 516 GSFOBK 521

## RESULT 7

US-09-866-108-3

Sequence 3, Application US/09866108  
Patent No. US2002004800A1

GENERAL INFORMATION:

APPLICANT: GU, Yizhong

APPLICANT: PENN, Sharon G.

APPLICANT: HANZEL, David K.

APPLICANT: CHEN, Wensheng

APPLICANT: SHANNON, Mark

TITLE OF INVENTION: MYOSIN-LIKE GENE EXPRESSED IN HUMAN HEART AND MUSCLE

FILE REFERENCE: AECOMICA-7

CURRENT APPLICATION NUMBER: US/09/866,108

PRIOR FILING DATE: 2001-05-25

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: GB 24263,6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 60/266,860

PRIOR FILING DATE: 2001-02-05

NUMBER OF SEQ ID NOS: 15752

SOFTWARE: Aecomica Sequence Listing Engine

SEQ ID NO: 3

LENGTH: 2568

TYPE: PRT

ORGANISM: Homo sapiens

US-09-866-108-3

Qy 29 VODEVLENTSSEAFIENLRRRRENLIYTYIGVLYSVNPFYRLOLIYSROHMERYGV 88  
Db 573 VEDLASLIS-VNESSVNLTLQRYKAOLHTCTGPDLLIOP-RKPSVSAKVKGRKD 630  
Qy 89 SFYEVPPHFAVADIVYRALRTERRDQAVMISGESGAGCTEATKRLQVATTCAPERG 148  
Db 631 G---LPAHIGSMAORAVYMLNORRDOSTVALGRSGAGATTCCEOVLELVGMAGVD-G 686  
Qy 149 GAVRDLLQSNVLEAFNAKTLRDNSSRFQKYADVOVDFKGAIVNGHILSYLLEKSYV 208  
Db 687 RVSVEKIRATFTVLARAFGSVMAHSRSATRFQVMSVSDFNATGRITTAQLOLMLEKSYV 746  
Qy 209 VQNHGERNHFVYQDLBEGDEETLRGLERNPOSYLYVGGCAKYSI-----N 260  
Db 747 AROPEGESNFIYVSOMLAGLIDLRLTELNLIH-----QMASSSFQMGVSKPE 794  
Qy 261 DK-----SDMKVVRKALSYDFTDEVEDLLSIVASVLIHGNHFADEDSNNOVTTENL 316  
Db 795 DKOKAAAFQAOLGAMEMIGISESEQRARVLAALVHIGAA-GACKYGRQFMRFEKA 852  
Qy 317 KYLTRLGVEGTTLREA-LTH--RKII-----AKGELLSPLNLEQAAVARDAL 362  
Db 853 NYAAELALGCFEYELNATATFKHHLROIQMTGPRMGJEDEFTSSGLKM-TGVDCVEGM 911  
Qy 363 AKAVYSRFTTWLVKRIINRSLSKDAESPSWRSTYVLGLDIYGFVEYQHN-----SFEQ 416  
Db 912 ASGLYQELFAAVVSLINRSFSHHLSMAS-----IMVDSFGQFNPHQKDRATFEE 965  
Qy 417 FCINVCNEKLODLFIETLLKSQDEYEAGIMPEVQYNNKIICDLYE----- 465  
Db 966 LCHNVAHERLQLLFYORFVSTLQRYOEGV--PVQF-----DLPPSPGTVAAYVD 1015  
Qy 466 -----EKFGIISILDEECLRPGEAATDLFLEKLEDTYVPRHPLTHKLA 510  
Db 1016 QNPQOQVRLPAGGQADARGLFWLDEEVHVEG-SDSVYLERLCACTEKKAGTEGSSA 1074  
Qy 511 DQTRKSLDRGEFRLHYAG-EVTVSYVFLDKNN-----DLFRNLKETMCS-- 557  
Db 1075 LRTCEQPL--QCEIFHQGMDPVRYDLTGWLHRAKPNLSALDAPVQLQSKREELRSLEF 1131  
Qy 558 -----SMNPT-----MACFDKSELSDKK-----RPEYATQFKASLLQVLEI 595  
Db 1132 QARAKLPYCAVAVGLEGTQALQSRMVRTFASLSAABRRKAPCSQIKQOMALITSM 1191  
Qy 596 LRSEKAPYIRCIKPNDA-----KOPGR-----GDEVLIIRHOVKYGLGM 633  
Db 1192 ITRSRHLFIHCLVPNPVYVESRSGQSPPPQGRDPRGAGPLADIPALVQIAGFHL 1251  
Qy 634 ENLAVRRAGFA-----YRRKYEA---FLQRYKSLCPETPMWAGRPQDGVAVLVRLH 682  
Db 1252 EALRLHRTGYADHMGILTRRQFOVLADAPLKKIMSTSGI-----DERKAVEBELLETL 1305  
Qy 683 GYKPEEYKMGRTKIFIRPKTLFATEDSLVRRQSLATK-----IOAAMRGFIHROKE--L 736  
Db 1306 DLEKKAVALVGHQVFLK-----AGVLSRLEKQREKLVOSIVLFGAACGFLSRQEFKRL 1360  
Qy 737 RYKR-SALCIOS-----MWR-----GLIGRRKAKRKAQAQOTIR-LI 772  
Db 1361 KIRRLAQOCTOKNNVAVFLAVKDPMMWOLIGSLQPLISATIGEQLRAKAEELITLRLKLE 1420  
Qy 773 RGFILRHSPRCPENAFFLDHVRASFLNLN-----RRQLPNNVLDTSWTPPPALREASEL 827  
Db 1421 KSEKILRNEIR--QNTDLESKLTADLSLDADRPFQGDVACQVLESRRARLRQAFRVOGL 1478  
Qy 828 LRELCKNNVVKYCRSISPENKQLOQKAVASEIRK--GKKDNTPO-----SVPRILTS 879  
Db 1479 KSK---HEQVOKKLDVYNKO-LEEAQOKTOLNDEBPDPTGAGDEQOMREDCAMENEFRL 1534  
Qy 880 TRLG-----TEISPR--VLQSLSEPTIYAVPVYKRGYKPPRROL----- 922  
Db 1535 KRLQOCEERLDELTRAKLEOKLGE--LQSA-----YD--CAKKMAHOLKCKCHHLTCD 1585  
Qy 923 LPPSAVVI-----VEDAKVVKORIDVYANLTGISV-----SSLSDSLFVL 960

Query Match 9.2% Score 497 DB 10:

Best Local Similarity 22.6% Pred. No. 3.3e-30: Length 2568:  
Matches 266; Conservative 194; Mismatches 439; Indels 280; Gaps 51;



Db 1586 LEPTCVLLENQSRNHELEKQKQKFDLQALQALGESVEEKGREKREYTOENTSVHEMELGOL 1645  
QY 961 HVGREDNKO-----KGDVYLOS DHIEFTLTALSAADY 994  
Db 1646 OOLKQKQEQEASQLKQOVEMLODHRKRELLSGPSLGENCV 1684

RESULT 8  
US-09-925-301-923

Sequence 923, Application US/09925301  
Patent No. US20020052308A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA106  
CURRENT APPLICATION NUMBER: US/09/925,301  
PRIOR FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05882  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
NUMBER OF SEQ ID NOS: 1694  
SOFTWARE: Patent Ver. 2.0  
SEQ ID NO 923  
LENGTH: 296  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-925-301-923

Query Match  
Best Local Similarity 4.0%; Score 215.5; DB 10; Length 296;  
Matches 80; Conservative 55; Mismatches 127; Indels 39; Gaps 12;

QY 763 WAAQTIRLRIRGFIHSPRCENAFELDHVARSFLNLRLQDLR-NVLDTSWPTPPAL 821  
Db 8 WGLGVRRREYKRFKFRANAGK-KIEFTLQIVQKYFLEMKRNMPSLSPIDKNMPSRYLF 66  
QY 822 REASLELRELCKKNWV---KYCRSISPEWKOLOQKAVASEIFKGMKNDYPOSYRLEF 878  
Db 67 LDSTH--KELKRIFHLWCKKRYRQFTDOOKLIYEKLEASLEFKDKKALYPSSVGQPFQ 124  
QY 879 STRLCTEIRSP---RVLOSIGSEPIQYAVPVVKYIDKGYKPPRQLLTTPSAVIVE-- 932  
Db 125 GAVL---EINKNPKYKLLKDAIEKIIIAEVVKNIRANKGSTRIFELTNLNNLLADQK 181  
QY 933 DAKKQKIDVANTLGISVSSLSLFLVHOR-EDNKOKGDVYLOS DHIEFTLT---TA 988  
Db 182 SGQIKSEVPADVTKYKSSQNDGFPAVHLKEGSEASAGDLEFSSDHLIEATKLYRTT 241  
QY 989 LSADRVN-NINI-----NQSITRAGCGPRGDCIIDFTSGSELLITKAKNGLAVV 1037  
Db 242 LSQTKKRLNIEISDFLVQGRDQKVCVFKIQGNKNGSVP-----TCRRKNRLLEV 293  
QY 1038 A 1038  
Db 294 A 294

RESULT 9  
US-09-925-299-937

Sequence 937, Application US/09925299  
Patent No. US20020055627A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA102  
CURRENT APPLICATION NUMBER: US/09/925,299  
PRIOR FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05883  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12

NUMBER OF SEQ ID NOS: 1556  
SOFTWARE: Patent Ver. 2.0  
SEQ ID NO 937  
LENGTH: 237  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (79)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
LOCATION: (85)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-299-937

Query Match  
Best Local Similarity 3.8%; Score 207.5; DB 10; Length 237;  
Matches 70; Conservative 32; Mismatches 87; Indels 49; Gaps 9;

QY 649 YEAFLORYKSLCEPTWPMAGRPDQAV--LVRLGYKPEEYKMGRTKIFIRPKTLFA 706  
Db 4 YEKFLHRYKIMISEFTWPN-HDLPSEKAVKRLIERCGFQ--DVAVGKTKIFIRPTLT 61  
QY 707 TENDLEVRQSLATKIQAMRGFHWKQFLKRVKSAICIOSMWRGTLGRKAKKRAA 766  
Db 62 LE---ELRAQMLI-----KIVFLQXWNGTAYARR-YKRTKAL 97  
QY 767 TIRRLRGFIHSPRCENAFELDHVARSFLNLRLQDLRNVLDTSWPTPPALREASE 826  
Db 98 TIRRYRYRYKVK-----SYHEVARRFHGVYTMNDYKQHV---KWPSPKVLRFEE 146  
QY 827 LLRELCKKNWV---KYCRSISPEWKOLOQKAVASEIFKGMKNDYPOSYRLEF 874  
Db 147 ALQITFNWRASQLIKSIPASDLPQVRKAYAAVEMLKGRADLGLORAWEGNYLASKP 204

RESULT 10  
US-09-864-761-34208

Sequence 761-34208 Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecm1ca-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
PRIOR FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663



```

; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 48341
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC010322.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.8
; OTHER INFORMATION: SWISSPROT HIT: P34092, EVALUATE 4.00e-14
; OTHER INFORMATION: EST_HUMAN HIT: BE513695.1, EVALUATE 5.00e-16
US-09-864-761-48341

```

```

Query Match          2.7%  Score 148; DB 10; Length 77;
Best Local Similarity 60.0%; Pred. No. 4.6e-05;
Matches 27; Conservative 7; Mismatches 11; Indels 0; Gaps 0;
QY 180 GKVDVDFGKAPVGHITLSTYLLKSRVYHOMHNGENRHFVFLQ 224
Db 13 GKVEIOLFSGRGERDGSKISNFIKLSRYVMQNMENRHFHYQV 57

```

```

RESULT 13
; Sequence 1, Application US/09803126
; Patent No. US20020099190A1
; GENERAL INFORMATION:
; APPLICANT: Brooks, Alan R.
; APPLICANT: Deng, Gary G.
; APPLICANT: Rubanyi, Gabor M.
; TITLE OF INVENTION: Estrogen-Regulated Unconventional Myosin-Related
; FILE REFERENCE: 01503-000310US
; CURRENT APPLICATION NUMBER: US/09/803,126
; PRIOR APPLICATION NUMBER: US 60/188,488
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1337
; TYPE: PRT
; ORGANISM: Mus sp.
; FEATURE:
; OTHER INFORMATION: mouse myosin related protein (MRP)
US-09-803-126-1

```

```

Query Match          2.7%  Score 145.5; DB 10; Length 1337;
Best Local Similarity 24.9%; Pred. No 0.0062;
Matches 52; Conservative 42; Mismatches 78; Indels 37; Gaps 8;
QY 563 MACQEDKSEL---SDKKRPETVAHQFMSLQLVEILRSKPAV-INCIRPNDAKQGRF 618
Db 1 MGSLEQEAEPQAGTEQNKP-TLASRFOOTGLDLARLSRGHYVHICLNPTGKIPGL 59
QY 619 DEVLIRHQVLYGLMENVRRVRRAGEAVRRKYEAFIQKSLCPETPMNAGRPQDGV--- 675
Db 60 DVGVHAGQLRQAGILLETIGRSTHFVYRVSFQVFLARFHAL-----GSGRQKASDOE 112
QY 676 ---AVLVRLHLYGKPEEKMGKRTKIFI-----RPFRTLFATEDSLLEV-----R 714
Db 113 RGSALILESEVLGAESPLVHLGVTLQEQGQQLQEDLMQRRSQALLTLHRLGRLACTIQ 172
QY 715 RGSALITQAMGCFHNRQKFLVYKRSAT 743
Db 173 RLRLPRMQARVRLQGLARKRYIQ-RRSAL 200

```

```

RESULT 14
US-09-864-761-39854
; Sequence 39854, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR APPLICATION NUMBER: 2001-05-23
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117

```

```
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 39854
;; LENGTH: 100
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC005323.1
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.3
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.6
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 19
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.4
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.5
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.4
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.1
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.1
;; OTHER INFORMATION: SWISSPROT HIT: P13535, EVALUE 2.00e-54
;; OTHER INFORMATION: EST_HUMAN HIT: F00185.1, EVALUE 7.00e-25
US-09-864-761-39854
```

```
Query Match 2.7%; Score 144.5; DB 10; Length 100;
Best Local Similarity 33.1%; Pred. No. 0.00013;
Matches 41; Conservative 15; Mismatches 29; Indels 39; Gaps 5;
```

```
OY 470 GIISLIDECCLRGATDTLTLEKLEDT-----VKPHHFLTKLADOKTRKSLDRG 521
DB 3 GIFSLIEECMP-KATDTSFKNKLYDHLGKSANFQKP-----KVVGKKA 49
OY 522 EFRLLHAGVTVYTGFLDKNNDLFRNLKETMSSNMPIMAGCFDSELSDKRPETV 581
DB 50 HFLHLHYAGVDYNTGWLKKNKDL-----PLNDYVGLYQXSAM-----KTL 91
OY 582 ATQF 585
DB 92 ASLF 95
```

```
RESULT 15
US-09-864-761-41762
;; Sequence 41762, Application US/09864761
;; Patent NO. US20020048763A1
;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharon G.
;; APPLICANT: Rank, David R.
;; APPLICANT: Hanzel, David K.
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; FILE REFERENCE: Acomica-X-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263,6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
```

```
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 41762
;; LENGTH: 102
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
```

```
;; OTHER INFORMATION: MAP TO AC005323.1
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.7
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.6
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
;; OTHER INFORMATION: EST_HUMAN HIT: F00185.1, EVALUE 5.00e-26
;; OTHER INFORMATION: SWISSPROT HIT: P13535, EVALUE 4.00e-50
US-09-864-761-41762
```

```
Query Match 2.7%; Score 144.5; DB 10; Length 102;
Best Local Similarity 34.5%; Pred. No. 0.00013;
Matches 39; Conservative 15; Mismatches 38; Indels 21; Gaps 3;
```

```
OY 470 GIISLIDECCLRGATDTLTLEKLEDT-----VKPHHFLTKLADOKTRKSLDRG 521
DB 3 GIFSLIEECMP-KATDTSFKNKLYDHLGKSANFQKP-----KVVGKKA 49
OY 522 EFRLLHAGVTVYTGFLDKNNDLFRNLKETMSSNMPIMAGCFDSELS 574
DB 50 HFLHLHYAGVDYNTGWLKKNKDLPLNETVGLYQXSAMKTIAOLFSGAOTAE 102
```

Search completed: December 9, 2002, 16:55:10  
Job time : 26 secs





```

QY 437 SEQEEYEAAGIAMEPVOYFNKKIICDLYBEKFKGIISLDEBCLRPGEATDLPTEKLED 496
DB 421 SEQEEYEAAGIAMEPVOYFNKKIICDLYBEKFKGIISLDEBCLRPGEATDLPTEKLED 480
QY 497 TVKPHHFLTHKLADQKTRKSLDGEFRLHYAGEVTVSYNGFLDKNNDDLFRNLKFTMC 556
DB 481 TVKHHHFLTHKLADQKTRKSLDGEFRLHYAGEVTVSYNGFLDKNNDDLFRNLKFTMC 540
QY 557 SSMNPIMACPFKSELSSKKRPETVATQPKMSLQVLEILRSKPEPAYIRCIKPNDAKOPG 616
DB 541 SSMNPIMACPFKSELSSKKRPETVATQPKMSLQVLEILRSKPEPAYIRCIKPNDAKOPG 600
QY 617 RDEVLIRHQQVYLGEMENLVRBAGFAVRKRYEAFQRYKSLCPEWPMWAGRPDQVA 676
DB 601 RDEVLIRHQQVYLGEMENLVRBAGFAVRKRYEAFQRYKSLCPEWPMWAGRPDQVA 660
QY 677 VLVRHLGYKPEEYKMGRTKIFIRPKTLFATEDSLLEVRSQSLATKIOAMRGFHMROKFL 736
DB 661 VLVRHLGYKPEEYKMGRTKIFIRPKTLFATEDSLLEVRSQSLATKIOAMRGFHMROKFL 720
QY 737 RVKRSALCIOSWMRGTLGRKRAKRRKMAQTIRLLRGFTLRHSPRCPENAFFLDHVRS 796
DB 721 RVKRSALCIOSWMRGTLGRKRAKRRKMAQTIRLLRGFTLRHSPRCPENAFFLDHVRS 780
QY 797 FLNLIRQLPRNVLDTSWTPPPALREASELLRELCMKMVKYCRSISPEKKOOLQOKA 856
DB 781 FLNLIRQLPRNVLDTSWTPPPALREASELLRELCMKMVKYCRSISPEKKOOLQOKA 840
QY 857 VASEIFKGGKNDYPOSVPRFLFSTRLGTEISPRVQSLGSEPIQYAVPVYKIDKKGYK 916
DB 841 VASEIFKGGKNDYPOSVPRFLFSTRLGTEISPRVQSLGSEPIQYAVPVYKIDKKGYK 900
QY 917 RPRQLLTPSAVYVEDAVKORIDYANLTGTSVSSLSDSLFLVHVOREDKOKGDVYLQ 976
DB 901 RSRQLLTPSAVYVEDAVKORIDYANLTGTSVSSLSDSLFLVHVOREDKOKGDVYLQ 960
QY 977 SDHVIETLTKTALSADRVNNININGSTTFAGGPGROGIDFTSGSELITTKAKNGHLAV 1036
DB 961 SDHVIETLTKTALSADRVNNININGSTTFAGGPGROGIDFTSGSELITTKAKNGHLAV 1020
QY 1037 VAPRLNSR 1044
DB 1021 VAPRLNSR 1028

RESULT 2
A:Species: Homo sapiens (man)
C:Date: 19-May-2000 #sequence_rev1sion 19-May-2000 #text_change 08-Sep-2000
C:Accession: A59253
Genomes: 40, 332-341, 1997
R:Cozet, F.; Amraoui, A.E.; Blanchard, S.; Lenoir, M.; Ripoll, C.; Vago, P.; Hamel, C.;
A:Title: Cloning of the genes encoding two murine and human cochlear unconventional type
A:Reference number: A59253; MUID:97237053; PMID:9119401
A:Accession: A59253
A:Status: Preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1028 <CRO>
A:Cross-references: GB:X98507; NID:91926310; PIDN:CAA67131.1; PID:91926311
A:Experimental source: dev stage adult; tissue type kidney
A:Genetics:
A:Gene: myo-1b
A:Map position: 17p3.2-p13.3
C:Superfamily: brush border myosin heavy chain I; myosin motor domain homology
F.14-683/Domain: myosin motor domain homology <NMO>

Query Match 95.3%; Score 5143; DB 2; Length 1028;
Best local similarity 96.2%; Pred. No. 0;
Matches 989; Conservative 23; Mismatches 16; Indels 0; Gaps 0;

QY 17 MESALTARDRYGVODFVLENTSEAFIENLRRFRRENTLYTYIGPVLSVNPYRLOI 76
I:|||||

```

```

DB 1 MSALTARDRYGVODFVLENTSEAFIENLRRFRRENTLYTYIGPVLSVNPYRLOI 60
QY 77 YSRQHMERYRGVSEYEPVPHLEAVADYVYRALRTERDQAVMISGEGACKTATRLLQ 136
DB 61 YSRQHMERYRGVSEYEPVPHLEAVADYVYRALRTERDQAVMISGEGACKTATRLLQ 120
QY 137 FYAETCPAPBEGGAVRDLQSNPYLEAFGNAKTLRNDNSRFGKYVDVDFDKGAPVGG 196
DB 121 FYAETCPAPBEGGAVRDLQSNPYLEAFGNAKTLRNDNSRFGKYVDVDFDKGAPVGG 180
QY 197 HILSYLEKSVYVHONGERNFHFVYQLLGGESETLRGLGERNPQSLYTVKGQCAVY 256
DB 181 HILSYLEKSVYVHONGERNFHFVYQLLGGESETLRGLGERNPQSLYTVKGQCAVY 240
QY 257 SSINDKSDMKVMKRAKLSVIDFTEDEVEDLSIYASVJHGNHFADEDSNNOVTTENQ 316
DB 241 SSINDKSDMKVMKRAKLSVIDFTEDEVEDLSIYASVJHGNHFADEDSNNOVTTENQ 300
QY 317 KYLTRLGEGGTLRREALTRKTIKAGEBELSPLEQAAVYRDLAKAVYSRTFTWLV 376
DB 301 KYLTRLGEGGTLRREALTRKTIKAGEBELSPLEQAAVYRDLAKAVYSRTFTWLV 360
QY 377 KIRNSLASKDAESPMSRSTTVGLADIYGEVFOHNSFPOFCINCNKLOQLTELTK 436
DB 361 KIRNSLASKDAESPMSRSTTVGLADIYGEVFOHNSFPOFCINCNKLOQLTELTK 420
QY 437 SEQEEYEAAGIAMEPVOYFNKKIICDLYBEKFKGIISLDEBCLRPGEATDLPTEKLED 496
DB 421 SEQEEYEAAGIAMEPVOYFNKKIICDLYBEKFKGIISLDEBCLRPGEATDLPTEKLED 480
QY 497 TVKPHHFLTHKLADQKTRKSLDGEFRLHYAGEVTVSYNGFLDKNNDDLFRNLKFTMC 556
DB 481 TVKHHHFLTHKLADQKTRKSLDGEFRLHYAGEVTVSYNGFLDKNNDDLFRNLKFTMC 540
QY 557 SSMNPIMACPFKSELSSKKRPETVATQPKMSLQVLEILRSKPEPAYIRCIKPNDAKOPG 616
DB 541 SSMNPIMACPFKSELSSKKRPETVATQPKMSLQVLEILRSKPEPAYIRCIKPNDAKOPG 600
QY 617 RDEVLIRHQQVYLGEMENLVRBAGFAVRKRYEAFQRYKSLCPEWPMWAGRPDQVA 676
DB 601 RDEVLIRHQQVYLGEMENLVRBAGFAVRKRYEAFQRYKSLCPEWPMWAGRPDQVA 660
QY 677 VLVRHLGYKPEEYKMGRTKIFIRPKTLFATEDSLLEVRSQSLATKIOAMRGFHMROKFL 736
DB 661 VLVRHLGYKPEEYKMGRTKIFIRPKTLFATEDSLLEVRSQSLATKIOAMRGFHMROKFL 720
QY 737 RVKRSALCIOSWMRGTLGRKRAKRRKMAQTIRLLRGFTLRHSPRCPENAFFLDHVRS 796
DB 721 RVKRSALCIOSWMRGTLGRKRAKRRKMAQTIRLLRGFTLRHSPRCPENAFFLDHVRS 780
QY 797 FLNLIRQLPRNVLDTSWTPPPALREASELLRELCMKMVKYCRSISPEKKOOLQOKA 856
DB 781 FLNLIRQLPRNVLDTSWTPPPALREASELLRELCMKMVKYCRSISPEKKOOLQOKA 840
QY 857 VASEIFKGGKNDYPOSVPRFLFSTRLGTEISPRVQSLGSEPIQYAVPVYKIDKKGYK 916
DB 841 VASEIFKGGKNDYPOSVPRFLFSTRLGTEISPRVQSLGSEPIQYAVPVYKIDKKGYK 900
QY 917 RPRQLLTPSAVYVEDAVKORIDYANLTGTSVSSLSDSLFLVHVOREDKOKGDVYLQ 976
DB 901 RSRQLLTPSAVYVEDAVKORIDYANLTGTSVSSLSDSLFLVHVOREDKOKGDVYLQ 960
QY 977 SDHVIETLTKTALSADRVNNININGSTTFAGGPGROGIDFTSGSELITTKAKNGHLAV 1036
DB 961 SDHVIETLTKTALSADRVNNININGSTTFAGGPGROGIDFTSGSELITTKAKNGHLAV 1020
QY 1037 VAPRLNSR 1044
DB 1021 VAPRLNSR 1028

RESULT 3
S41749

```

myosin heavy chain I beta - bovine  
 N:Alternate names: myosin IB heavy chain  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 20-May-1994 #sequence\_revision 02-Jun-1995 #text\_change 02-Feb-2001  
 C:Accession: S41749; A55744; S33497  
 R:Zhu, T.; Ikebe, M.  
 FEBS Lett. 339, 31-36, 1994  
 A>Title: A novel myosin I from bovine adrenal gland.  
 A:Reference number: S41749; MUID:94148088; PMID:8313976  
 A:Accession: S41749  
 A:Molecule type: mRNA  
 A:Residues: 1-1028 <ZHU>  
 A:Cross-references: GB:U03420; NID:g436936; PIDN:AAA17565.1; PID:g436937  
 A:Experimental source: adrenal gland  
 R:Reizes, O.; Barylko, B.; Li, C.; Suedhof, T.C.; Albanesi, J.P.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 6349-6353, 1994  
 A>Title: Domain structure of a mammalian myosin Ibeta.  
 A:Reference number: A55744; MUID:94294379; PMID:8022785  
 A:Accession: A55744  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-6,'C','8'-96,'R','98-221,'ECT',223-281,'N',283-296,'GTSSST',303,'PGPWVR',307-721,'L',723-909,'WSF',913,916-988,'SSQ',992-994,'AGMASTLSH',1004,'ARSCSSPREPT',1016,'TW  
 A:Cross-references: EMBL:222852  
 A:Experimental source: brain  
 C:Superfamily: brush border myosin heavy chain I; myosin motor domain homology  
 C:Keywords: actin binding; ATP; calmodulin binding; nucleotide binding; P-loop  
 F:14-683/Domain: myosin motor domain homology <MMOT>  
 F:105-112/Region: nucleotide-binding motif A (P-loop)  
 F:576-595/Region: actin binding #status predicted  
 F:699-743/Region: calmodulin binding #status predicted  
 F:111/Binding site: ATP (Lys) #status predicted

Query Match 95.0%; Score 5128; DB 2; Length 1028;  
 Best Local Similarity 95.6%; Pred. No. 0;  
 Matches 983; Conservative 25; Mismatches 20; Indels 0; Gaps 0;

QY 17 MESALTARRGVGVDFVLENTSEAFTEENRRRENLITTYIGPVLSVNPYRDLOI 76  
 DB 1 MESALTARRGVGVDFVLENTSEAFTEENRRRENLITTYIGPVLSVNPYRDLOI 60  
 QY 77 YRQHMERGRGVSEFVPHLFAVADTVRALRTERDDQAVMISGSGAGTEATKRLIQ 136  
 DB 61 YRQHMERGRGVSEFVPHLFAVADTVRALRTERDDQAVMISGSGAGTEATKRLIQ 120  
 QY 137 FYAETCPABERGAGVARDLLQSNPVLEAFGNKATLRNDSSRFQKMDVDFPKGAPVG 196  
 DB 121 FYAETCPABERGAGVARDLLQSNPVLEAFGNKATLRNDSSRFQKMDVDFPKGAPVG 180  
 QY 197 HILSYLLEKSRVYVHONGHGRNFHVYQLLEGGEDETLRLGLERNPQSLTYLVKGCACAV 256  
 DB 181 HILSYLLEKSRVYVHONGHGRNFHVYQLLEGGEDETLRLGLERNPQSLTYLVKGCACAV 240  
 QY 257 SSINRSDMKVVRKALSVDFTEDEVEDLSTIVASYLHGNTHFADEDSNNQVTTENQI 316  
 DB 241 SSINRSDMKVVRKALSVDFTEDEVEDLSTIVASYLHGNTHFADEDSNNQVTTENQI 300  
 QY 317 KYLVRLLGEGTTLREALTHRKIIAKGELLSPNLNQAAVARDALAKAVSRTEFTWLVR 376  
 DB 301 KYLVRLLGEGTTLREALTHRKIIAKGELLSPNLNQAAVARDALAKAVSRTEFTWLVA 360  
 QY 377 KINRSLASDASPSWRSTTVYGLDLYGFVFNHNSFQPCINCNKRLQDLFTLTLK 436  
 DB 361 KINRSLASDASPSWRSTTVYGLDLYGFVFNHNSFQPCINCNKRLQDLFTLTLK 420

QY 437 SQEEYEAEGLMEPEVOYFNKKIICDLYEKKGIISILDECLRGAEATDITPLEKLD 496  
 DB 421 SQEEYEAEGLMEPEVOYFNKKIICDLYEKKGIISILDECLRGAEATDITPLEKLD 480  
 QY 497 TVKPPHPLTHKLAQOKPKKSLDRGEFRLHYAGVTSVTPFLKNNDLFRNLKETMC 556  
 DB 481 TVKPPHPLTHKLAQOKPKKSLDRGEFRLHYAGVTSVTPFLKNNDLFRNLKETMC 540  
 QY 557 SSANPIMACFPKSELSDKRPETVATOPKMSLLOVELLRKPEYAVICIRPNDAKOPG 616  
 DB 541 SSANPIMACFPKSELSDKRPETVATOPKMSLLOVELLRKPEYAVICIRPNDAKOPG 600  
 QY 617 RDEVELIRHGVYKIGLMENLRVRAGFAVRRKYEAFLQRYKSLCEPTWDMAGRPDQVA 676  
 DB 601 RDEVELIRHGVYKIGLMENLRVRAGFAVRRKYEAFLQRYKSLCEPTWDMAGRPDQVA 660  
 QY 677 VLVRLHGYKPEEYKMGKRTKIFTRFKTLFATDSLEVRQSLATKIQAAMRGFHWKQFL 736  
 DB 661 VLVRLHGYKPEEYKMGKRTKIFTRFKTLFATDSLEVRQSLATKIQAAMRGFHWKQFL 720  
 QY 737 RKRSATCIGSMMRGTLGRKAKRMAQOTLRRLRGFTLRHSPRCPEANAFLOHVRAS 796  
 DB 721 RKRSATCIGSMMRGTLGRKAKRMAQOTLRRLRGFTLRHSPRCPEANAFLOHVRAS 780  
 QY 797 FLNLRLROLPRNVLDTSWTPPPALREASELLRELCMKMMWKYCSISPEKKOQLOQKA 856  
 DB 781 FLNLRLROLPRNVLDTSWTPPPALREASELLRELCMKMMWKYCSISPEKKOQLOQKA 840  
 QY 857 VASEIFKGGKNDYPOSVPLRTISTRLGTEISPRVLOSISEPTIOYAVPVYKRRGYKP 916  
 DB 841 VASEIFKGGKNDYPOSVPLRTISTRLGTEISPRVLOSISEPTIOYAVPVYKRRGYKP 900  
 QY 917 RPRQLLPASAVYVEDAVKORIDVANTLGISVSSLSQSLVLYHQRDNKGQGVVLO 976  
 DB 901 RPRQLLPASAVYVEDAVKORIDVANTLGISVSSLSQSLVLYHQRDNKGQGVVLO 960  
 QY 977 SDHVLETLTKTALSADRVNNININOSITFAGPGKGDITDTSGESELLITRAKNHGLAV 1036  
 DB 961 SDHVLETLTKTALSADRVNNININOSITFAGPGKGDITDTSGESELLITRAKNHGLAV 1020

QY 1037 VAPRLNSR 1044  
 DB 1021 VAPRLNSR 1028

RESULT 4  
 myosin I beta - bullfrog  
 C:Species: Rana catesbeiana (bullfrog)  
 C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 02-Feb-2001  
 C:Accession: I51173  
 R:Metcalfe, A.B.; Chelliah, Y.; Hudspeeth, A.J.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 11821-11825, 1994  
 A>Title: Molecular cloning of a myosin I beta isoform that may mediate adaptation by  
 A:Reference number: I51173; MUID:95083594; PMID:7991542  
 A:Accession: I51173  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1028 <MET>  
 A:Cross-references: EMBL:U14549; NID:g602137; PIDN:AAA57192.1; PID:g602138  
 C:Genetics:  
 A:Gene: AMI beta  
 A:Introns: 639/72  
 C:Superfamily: brush border myosin heavy chain I; myosin motor domain homology  
 C:Keywords: nucleotide binding; P-loop  
 F:14-683/Domain: myosin motor domain homology <MMOT>  
 F:105-112/Region: nucleotide-binding motif A (P-loop)

Query Match 81.0%; Score 4371; DB 2; Length 1028;  
 Best Local Similarity 79.2%; Pred. No. 7,8e-272;  
 Matches 814; Conservative 112; Mismatches 102; Indels 0; Gaps 0;



QY	17	MSALTDARDVGVODEVLLLENFTNSEAFIENLRBRREFENLITYTGPVLVSNVPRJDOJ	76
Db	1	MSALTDARDVGVODEVLLLENFTNSEAFIENLRBRREFENLITYTGPVLVSNVPRJDOJ	60
QY	77	YSROJMEYRQCVSEYEPYEPHLEFADVYVYALTBTERDDVAWISGSGAGKTEATKRLQJ	13
Db	61	YSKQJMEYRQCVSEYEPYEPHLEFADVYVYALTBTERDDVAWISGSGAGKTEATKRLQJ	120
QY	137	FYAEICAPERBEGCAVRDRLQSNPVLAEAFGNATLTNNDMSRFGKYMDQDFPKCAPYVG	190
Db	121	YYAATCPVSDQVETVKKDRLLQSNPVLAEAFGNATLTNNDMSRFGKYMDQDFPKCAPYVG	180
QY	197	HLSVLYLEKSVYHONNGENRPHVFPQJLEGGEEETRLRGLGERNPQSTLYLVKQCAKV	25
Db	181	HILWVLYLEKSVYHONNGENRPHVFPQJLEGGEEETRLRGLGERNPQSTLYLVKQCAKV	240
QY	257	SSINDKDMKVMKRAKALSVIDTEDEVEDLLSTIVASVLYLGNTHFADEDSNAQVTTENQJ	310
Db	241	SSINDKDMKVMKRAKALSVIDTEDEVEDLLSTIVASVLYLGNTHFADEDSNAQVTTENQJ	300
QY	317	KYLTRLVLEGVTLRREALTRKTLIAAGEBLSPLNLEQAAVYAROLAALAAVYSRFFTVLVR	370
Db	301	KYLARLTSVSTVLRRESLTHKTKIAKEBLSPLNLEQAAVYAROLAALAAVYSRFFTVLVR	360
QY	377	KIRSLASKAESBRSKSTTVLGLDIDYGEVEYQHNSEFOCINCNKELDQJLEIETLK	430
Db	361	KIRKSLASKAETDHIKLGASVIGLIDYGEVEYQHNSEFOCINCNKELDQJLEIETLK	420
QY	437	SEOEYEAEAGIAMEPQYQYNNKTIIDOLYEBKKGITSLDECCIRPEAGDPTLFLEKLED	490
Db	421	SEODEYESEGIAEMEPQYQYNNKTIIDOLYEBKKGITSLDECCIRPEAGDPTLFLEKLED	480
QY	497	TYVPHPHFLTHKLADQTKRSLDRBEPRLHYAGEVITYVTGFLDKNNDLFLRNLKETWC	550
Db	481	TYVNNHPEHYTHKRGDQTKRKYVGRBEPRLHYAGEVINYVAGFLDKNNDLFLRNLKEVWC	540
QY	557	SSNMPIMAOCPFKSELSDKRPBETVATQFKMSLLOLVEYILRSKPEAVYIRCIKPNDAKQJG	610
Db	541	DSGNPPIAHQCFNNSSETLQKRPETAHQFKMSLSKMEILMSKQPSVYCIKPNDAKQJG	600
QY	617	RPEEVLIRHQVYVYLMEMLRYKRRAGFYRRKYAFVQRRKSLCPETPMWAGRPQDGYA	670
Db	601	RDEEVLIRHQVYVYLMEMLRYKRRAGFYRRKYAFVQRRKSLCPETPMWAGRPQDGYA	660
QY	677	VLVRLHGYKPEEYKMGRTKIEIRPEPTLFATDEDELEVRQSLKIOAARGHMQRKTL	730
Db	661	VLVKSLGYKPEEYKMGRTKIEIRPEPTLFATDEDELEVRQSLKIOAARGHMQRKTL	720
QY	737	RKRSRGLCQSNMROTLGRKRAAKRKAQAOTIRRLNGFTLHRSPPRCPENAFELDHVRAS	790
Db	721	HKHSVAVELOQSNMROTLGRKRAAKRKAQAOTIRRLNGFTLHRSPPRCPENAFELDHVRAS	780
QY	797	FLNLRLQJLRPNVLDTSWTPPPLRBASELRELECKNNVMKYGCSISPEBKQJLOOKA	850
Db	781	FLMLTYLRMQPSVLDKSWPVPPSLRASSELLECKNNVMKYGCSISPEBKQJLOOKA	840
QY	857	VASELFEKKKNNYQSPVPRLFISTRLCTEBISPRVLOSLGSEPIQYAVVYVYKDRKGYP	910
Db	841	VASELFEKKKNNYQSPVPRLFISTRLCTEBISPRVLOSLGSEPIQYAVVYVYKDRKGYP	900
QY	917	RPROLLTPSAVYDEAKQKQRIADVANTLGISVSSLSLQSLVLYHVPQDNQOKDQVLO	970
Db	901	RRROLLTPQNAAYVLEAKKKQKQRIADVANTLGISVSSLSLQSLVLYHVPQDNQOKDQVLO	960
QY	977	SDHYIELTLTKALSADVYNNININQOSTIFPAGCPGHDGJIDFTSGSELLTKAKNGHLAV	1030
Db	961	SDHYIELTLTKALTAEEINNININQOSTIFPAGCPGHDGJIDFTSGSELLTKAKNGHLAV	1020
QY	1037	VAPRLNSR 1044	
Db	1021	VAPRLNSR 1028	

RESULT 5  
 H75634  
 myosin-Ic - mouse (Fragment)  
 N:Alternate names: myosin-I beta  
 C:Species: Mus musculus (house mouse)  
 C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 08-Sep-2000  
 C:Accession: H75634  
 R:Cross: F.; Ambouli, A.E.; Blanchard, S.; Lenoir, M.; Ripoll, C.; Vago, P.; Hamel,  
 Genomics 40, 332-341, 1997  
 A:Title: Cloning of the genes encoding two murine and human cochlear unconventional t  
 A:Reference number: A59253; MUID:97237053; PMID:9119401  
 A:Accession: H75634  
 A:status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-807 <CRO>  
 A:Cross-references: GB:X96938; NID:g1924960; PIDN:CA67956.1; PID:g1924961  
 A:Experimental source: strain BALB/c; tissue type cochlea; dev stage adult  
 C:Genetics:  
 A:Gene: MGI:Myo1c  
 A:Cross-references: MGI:106612  
 A:Map position: 11:44.1  
 C:Superfamily: brush border myosin heavy chain I; myosin motor domain homology  
 F1:14-683/Domain: myosin motor domain homology (MW)

Query Match	77.28;	Score 4167;	DB 2;	Length 807;
Best Local Similarity	99.13;	Pred. No. 6.3e-259;		
Matches 800;	Conservative	3;	Mismatches 4;	Indels 0;
				Gaps 0;
Qy	17	MSALTAADRVGVDFVLENTFSEAFINLRRRPENITITITIPVLSVNPRILOI	76	
Db	1	MSALTAADRVGVDFVLENTFSEAFINLNGCRENITITITIPVLSVNPRILOI	60	
Qy	77	YSRQHMERYKVSLEYEPPHLFAVDITVRALFTERRDOAVMISGSGAGTEATKRLQ	136	
Db	61	YRRHMEKRYKVSLEYEPPHLFAVDITVRALFTERRDOAVMISGSGAGTEATKRLQ	120	
Qy	137	FYATCTCAPRGCAVNRDLQSNPVLEAFONAKTLRNDNSFRFGKMDVOPFKAPGG	196	
Db	121	FYATCTCAPRGCAVNRDLQSNPVLEAFONAKTLRNDNSFRFGKMDVOPFKAPGG	180	
Qy	197	HIISYLLERKRVYVHNGERNFHYVQLLEGBEETLRLLGLERNPOSYLYLVKQCAKV	256	
Db	181	HIISYLLERKRVYVHNGERNFHYVQLLEGBEETLRLLGLERNPOSYLYLVKQCAKV	240	
Qy	257	SSINDKSDMKVMKALSVIDFTEDEVEDLSTVASYLHNGTHFPAADSDNAQVTTENOL	316	
Db	241	SSINDKSDMKVMKALSVIDFTEDEVEDLSTVASYLHNGTHFPAADSDNAQVTTENOL	300	
Qy	317	KYLTRLGVEGTTLRALTHRRKIIAKGEELLSPNLLEQAAVARDALAKAVYSRTFTMLVR	376	
Db	301	KYLTRLGVEGTTLRALTHRRKIIAKGEELLSPNLLEQAAVARDALAKAVYSRTFTMLVR	360	
Qy	377	KINRSIASDAESPMSRSTTVGLLDIYGEFEQINSEPOCIYCNCKLQOLFELTLK	436	
Db	361	KINRSIASDAESPMSRSTTVGLLDIYGEFEQINSEPOCIYCNCKLQOLFELTLK	420	
Qy	437	SQCEYEAGIAMEVOYFNNKIIDOLVEEKFKGIISLDECLCPGATDLPFLEKLED	496	
Db	421	SQCEYEAGIAMEVOYFNNKIIDOLVEEKFKGIISLDECLCPGATDLPFLEKLED	480	
Qy	497	TYKPRHNFETHLAOQTRKSLDGEFFLLHAGEVITYSVYGFGLDKNNDLFLRLKETWC	556	
Db	481	TYKPRHNFETHLAOQTRKSLDGEFFLLHAGEVITYSVYGFGLDKNNDLFLRLKETWC	540	
Qy	557	SSMNPIMAOCPDKSELSDKKRPETVAQFKMSLQOLVEILRSKPEAPYIRCIRPNDAKOPG	616	
Db	541	SSMNPIMAOCPDKSELSDKKRPETVAQFKMSLQOLVEILRSKPEAPYIRCIRPNDAKOPG	600	
Qy	617	RDEVLILRHQVKYIGIMENLVRBAGFVRRKYAFAFQIRKSLCPEITPMAGRPDGVAA	676	
Db	601	RDEVLILRHQVKYIGIMENLVRBAGFVRRKYAFAFQIRKSLCPEITPMAGRPDGVAA	660	
Qy	677	VLVRLHGYKPEEYMGRIKTIIRPKPLFAETESLEVRQSLSTTKIOAARGGHMOKFEL	736	

|||||  
Db 661 VLVRLGKPEEYKMGKRIKIFIRPKTLFATEDSLVRRSLATKIQAAWGFHWKQKFL 720  
Oy 737 RKRSATCIQSWMNGTIGRRKAARKMAAOTIRRLIRGEIILRHSPRCENMFFLDHVAS 796  
Db 721 RKRSATCIQSWMNGTIGRRKAARKMAAOTIRRLIRGEIILRHSPRCENMFFLDHVAS 780  
Oy 797 FLNLRLRQLPNNVLDTSWTPPPALRE 823  
Db 781 FLNLRLRQLPNNVLDTSWTPPPALRD 807  
RESULT 6  
S45574  
myosin IB - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 02-Feb-2001  
C:Accession: S45574  
R:Strom-Morgan, N.; Skovronsky, D.M.; Artavanis-Tsakonas, S.; Mooseker, M.S.  
J. Mol. Biol. 239, 347-356, 1994  
A:Title: The molecular cloning and characterization of Drosophila melanogaster myosin-IB  
A:Reference number: S45573; MUID:94260541; PMID:8201616  
A:Accession: S45574  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1026 <STR>  
A:Cross-references: GB:U07596; NID:9466257; PIDN:AA19591.1; PID:9466258  
C:Superfamily: brush border myosin heavy chain I; myosin motor domain homology  
C:Keywords: nucleotide binding, P-loop  
F:14-681/Domain: myosin motor domain homology <MOT>  
F:105-112/Region: nucleotide-binding motif A (P-loop)  
Query Match 48.8%; Score 2632; DB 2; Length 1026;  
Best Local Similarity 51.5%; Pred. No. 2,1e-160;  
Matches 533; Conservative 177; Mismatches 300; Indels 24; Gaps 8;  
Oy 17 MESALTARDRVGVDFVLENTSEAFIENLRFRRENLIYTYIGVAVSVNPPYRDOI 76  
Db 1 METGHERDRAGVDFVLENYQSEAFIGNLKRFEEDLIYTYIGVAVSVNPPYKOLPI 60  
Oy 77 YSRQMEKRGVSFEYVPHLFAVADIVYRALRTRRQDAWISSESGAKTEARKRLQ 136  
Db 61 YTDNIVKAYRKHFEMPHLFAVDNMFRLIEENRCQVLISESGSGKTEASKKVLQ 120  
Oy 137 FYAETCPAPERGAARDLQSNPYLEAFNAKTLRNNSSEFGYVMPVQDFKCAPRG 196  
Db 121 FIACSGNQITVEGKDKLLSNPYLEAFNAKTLRNNSSEFGYVMPVQDFKCAPRG 180  
Oy 197 HILSYLLEKSRVYVQNHGERNFHFYQLLEGGEETLRRLGLERNPOSYLYLVKQCAKY 256  
Db 181 NILNLYLEKSRVYVQNHGERNFHFYQLLEGGEETLRRLGLERNPOSYLYLVKQCAKY 240  
Oy 257 SSINDKSMKVMKRLSYIDTFEDEVEDLSIVASVHLGNHFAADEDSNAQVTTENQ 316  
Db 241 TSINDASFQVQVQLVYIDFGEKEQREIFGIVASILHGVNF-TEVEGNAKVNRSQ 299  
Oy 317 KYLFLEGEETTLREALTHRRKIKAQSEELSPNLLEQAAVARDALAKAVSRPTMLVR 376  
Db 300 VTAARLLGVNASELEALTHRTIARQGVVSPNLQELAIYARQALAKAVYDRFLFSWLVQ 359  
Oy 377 KINRSLSKADSPSRWSTYVGLDIDYGFVEYFQHNSEFQPCINCNKLOOLFTELTLK 436  
Db 360 RLNLSLQAKETRA--SHNNWGIIDIDYGFVEYFQHNSEFQPCINCNKLOOLFTELTLK 416  
Oy 437 SEQFEYKAGLAMPBVQYFNKKITCDLVEERFKGITSLDEECRLRGVATDLTFLEKLED 496  
Db 417 SEQDYEKREGLEWLFVEYFDNKNVLCNLEEKNGKITSLDEECRLRGVATDLTFLEKLTQ 476  
Oy 497 TVKPRHFLTKLADOKRSLDRGEFRLLYAGEVYVYSGFLDKNNDLFRNLKETMC 556  
Db 477 KLAGNHNHVCHEKARANTIKMLRDEFRLYNHAGEVYVYSGFLDKNNDLFRDLKETLS 536  
Oy 557 SSMPNIMAQCFDKSELQKRPETVATQFKMSLLQVLVELNSKEPAYIRCIKPNDAKQ 616

: : : : :  
Db 537 KAGNIVASCEPEKELRSLKRPETAITQFASLNMIMLIMCKEPSYIRCIKPNLOQTAN 596  
Oy 617 RPEDELINHOVKYIGLMLNLRVRAGFAVRKYEAFLORYSLCEPTPMW--AGRPODG 674  
Db 597 VFNDLVLHQQVYIGLMLNLRVRAGFAVRKYEAFLORYSLCEPTPMW--AGRPODG 656  
Oy 675 VAVLRHNGYKPEEYKMGKRIKIFIRPKTLFATEDSLVRRSLATKIQAAWGFHWKQK 734  
Db 657 VQQLVKDGMDEEKRYVETKLFITWPRTLPTDEAYGEKKEIAITIQAHKGLMQRK 716  
Oy 735 FLVYKRSATCIQSWMNGTIGRRKAARKMAAOTIRRLIRGEIILRHSPRCENMFFLDHVR 794  
Db 717 YLKRQAVIIMQSYCRRLAQOAAKRRREADKIRAFIKGFITRDADPAGFNEEFLANAK 776  
Oy 795 ASFLNLRLRQLPNNVLDTSWTPPPALREASLLELCMKNNVMKYCCSISPEMKQOQLQ 854  
Db 777 RWMLLRLAKELPTKVLDSWPPAPGCEASGILHRLHRLHRLARLYRLKLPQQRQFEL 836  
Oy 855 KAVASEIFKGNKQNPQGVPPRLFISTRGTEI---SPRVLOSLSEPIQYAVPVYKXDR 911  
Db 837 KYLAKVYKGNKNNVASSVTFQEDRIPEKHIIQVNDPFASTPESDQKYSFCTKEDR 896  
Oy 912 KGYKPRPQLLTSPSAVYVEDAKYQ-----RIDYANLTGISVSSLSDSLFLYLVQR 964  
Db 897 HGKSRDFEILLSNKAIYVLDGKYTKOKHRLPLDKIDF-----TLTNHNDLAWYIRPL 950  
Oy 965 EDNKKQGVNLQSDHVIETLTKTALSADRVNNIN--NGSITFAGCPGRDGIIDTSSSE 1023  
Db 951 DKKKQGDILLIPRIETSYIIDTVGTAIVSLIDRNSLEHNHYKGGGVIDIQTGAEL 1010  
Oy 1024 LLITKAKNGHLAV 1037  
Db 1011 PGVARDK-GHLVIT 1023  
RESULT 7  
B45438  
myosin I beta, MMI beta - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Jan-2000  
C:Accession: B45438  
R:Sherr, E.H.; Joyce, M.P.; Greene, L.A.  
J. Cell Biol. 120, 1405-1416, 1993  
A:Title: Mammalian myosin I alpha, I beta, and I gamma: new widely expressed genes of  
A:Reference number: B45438; MUID:93194946; PMID:8449986  
A:Accession: B45438  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1-448 <SHB>  
A:Note: sequence extracted from NCBI backbone (NCBIP:131911)  
C:Superfamily: brush border myosin heavy chain I; myosin motor domain homology  
F:1-448/Domain: myosin motor domain homology (fragment) <MOT>  
Query Match 41.7%; Score 2252; DB 2; Length 448;  
Best Local Similarity 97.1%; Pred. No. 1.3e-136;  
Matches 435; Conservative 6; Mismatches 7; Indels 0; Gaps 0;  
Oy 152 LEAFGNATLRDNRSSRGKYMDVQDFKGAIVGGHILSYLLEKSRVYVQNHGERNFHF 221  
Db 1 LEAFGNATLRDNRSSRGKYMDVQDFKGAIVGGHILSYLLEKSRVYVQNHGERNFHF 60  
Oy 222 YOLLEGGEETLRRLGLERNPOSYLYLVKQCAKYSINDKSDMKLLKRLASVIDTFDE 281  
Db 61 YOLLEGGEETLRRLGLERNPOSYLYLVKQCAKYSINDKSDMKLLKRLASVIDTFDE 120  
Oy 282 VEDLLSIASVHLGNHFAADEDSNAQVTTENQKYLTRLGVEGTLREALTHRRKIIA 341  
Db 121 VEDLLSIASVHLGNHFAADEDSNAQVTTENQKYLTRLGVEGTLREALTHRRKIIA 180  
Oy 342 KGEELLSPNLLEQAAVARDALAKAVYSRFTYLVKIKINRSLSKADSPSRWSTYVGLL 401  
Db 181 KGEELLSPNLLEQAAVARDALAKAVYSRFTYLVKIKINRSLSKADSPSRWSTYVGLL 240

QY	402	DIGGEVQHNHSFQFCINVCNEKMLQOLFTELTJLKSQEEYEAEAGIAMEVQYFNNKIIIC	461
Db	241	DIYGEVQHNHSFQFCINVCNEKMLRLOFTLELTJLKSQEEYEAEAGIAMEVQYFNNKIIIC	300
QY	462	DLYEEKFKGIISILDEECRLRGEATDTLPLEKLEDDYVKPRPHFLVTHKLAQKTRKSLDRG	521
Db	301	DLYEEKFKGIISILDEECRLRGEATDTLPLEKLEDDYTKHHPHFLVTHKLAQKTRKSLDRG	360
QY	522	EFRLIHYAGEVTVSVTGFGLDKNNDLFRNLKETMCSMNPIMAQCFDKSELSDKRRETV	581
Db	361	EFRLIHYAGEVTVSVTGFGLDKNNDLFRNLKETMCSSTNPIMAQCFDKSELSDKRRETV	420
QY	582	ATQFKMSLLQVETLRSKEPARYTRCIKRP	609
Db	421	ATQFKMSLLQVETLRSKEPARYTRCIKRP	448

## RESULT 8

myosin-I, myric (alternatively spliced) - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 02-Feb-2001  
 C:Accession: BA5439  
 R:Ruppert, C.; Kroschewski, R.; Bahler, M.  
 J. Cell Biol. 120, 1393-1403, 1993  
 A:Title: Identification, characterization and cloning of myr 1, a mammalian myosin-I  
 A:Reference number: BA5439; MUID:93194945; PMID:8445985  
 A:Accession: BA5439  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1078 <RP>  
 A:Note: sequence extracted from NCBI backbone (NCBIN:128956, NCBIPI:128986)  
 C:Superfamily: brush border myosin heavy chain I; myosin motor domain homology  
 C:Keywords: nucleotide binding; P-loop  
 F:18-688/Domain: myosin motor domain homology <MMOT>  
 F:108-115/Region: nucleotide-binding motif A (P-loop)

## Query Match

Query Match	39.1%;	Score 2109;	DB 2;	Length 1078;
Best Local Similarity	42.8%;	Pred. No. 7.5e-127;		
Matches 463;	Conservative 183;	Mismatches 346;	Indels 90;	Gaps 19;

```

QY 27 VGVQDVELEENFNSAFLIENRRPRELLYTTYIGPVLVSANPYRQLOIYSQHIMERYR 86
Db 15 IGVGDVILVEEPL-NEETFTIDNKKRKHHEIYTYIGSVASISVNPYRSJPIISPEKVEEDY 73
QY 87 GVSFTVEVPHLFAVDJYRALTERRDOAVMISGESSAGKTEATKRLJOYAEYTCAPAE 146
Db 74 NRRFELSPHILALSDAEVRSRLDDQOCILITGESAGKTEASKLWMSVAVAVCKGA 133
QY 147 RGGAVBDRLLJOENPLVEAFNANKTLRNSSRFQYMDYQDFEGAVYGGHILSYLLEKS 206
Db 134 EVNQVKOEOLTPYLEAFNANKYIRANDSSRFQYMDIEFDEGDPJLGGVSIYVLEKS 193
QY 207 RVVHNHGERNFHYVYQLLEGGEEBTLRLGLEERNPOSYLYLVKGQCACVSSINDKDMK 266
Db 194 RVVKQPRERNFHYVYQLLSGASEELHKLTERDQFRNYVL-SLDSAKWGVDDANFR 252
QY 267 VRRKALSYIDFEDVEDVELLSIASVYLHGIHHAEDDS- - -AQVTTENOLKUTJRL 322
Db 253 TVRNMQIVGSDPEABESYLEVAVAVLKLTGITEKPRFSRNMGIDESKIKXKNELKEICEL 312
QY 323 LGVEETTLREALVHRKIIAKGEELSLPNTLEQAAVYARDALAKAVASTFTWVLRKINRSL 382
Db 313 TSLDVLVERAFSFRTEVAKOEKSTLVAOVAAYARDALAKMLYRSLTSMVLYNRINEEI 372
QY 383 ASKDAESPMSWSTVVLGILDIYGEVYQHNSFEQFCINCNENKLOQLEFTELKSFQEY 442
Db 373 KAQTKVRKK- - -VMGVLDIYGELEFEDMSFQFIINCNEKLOQIFLETLKEQEY 427
QY 443 EAESIAMEPVOYFNKKITCDIVEEKFKFGISIDEECLRGEGATDITFLLEKIEDYKPH 502
Db 428 IREDIEWTHIDIFYFNNAITCDILENNNTGLAMIDECLRGVYTDITFLLEKLNQVCATHQ 487

```

Qy	503	HLFH---KLADQKTRKSLDRGFERLLHNGEYTTYSVTFGLKNDNLLFLRLKETMSSMN	560
Db	488	HFSRMSKSRFLNDDTLTSHSCRFJONHAGVKLYOVEGFVRKNDNLLYRDLISQAMWACH	547
Qy	561	PIMAOCDSELS--DKRPETVAQOFKMSLJOLVETLRSKEPAPICITPENOAKORGR	618
Db	548	ALJXSLFPEBGNPAKVLKRPPTAGSOFAVSATVLMKLNJQTKNPNYINCILPNOKKAHIF	607
Qy	619	DEVLIHROKUYGLMENLHVBRAGAVYRKKEAEALQKSLCEPTWIMAGRPQGVAVL	678
Db	608	SESLVCHQJRIYIGLELNVBRAGAGFAQAVBPCLERKMKJCKQTMHMGPARSGEVL	667
Qy	679	VRLJLCKPPEYKMGRTKJFIREPKTLFATEDSELEVRRQSLATKJQAMRGFHRQKFLRV	738
Db	668	FNLEJLPEVEYSFGRSKJFIRNRPJTLFOEDLRKORLEDLATLQKLYRGWCKRTHFLM	727
Qy	739	KRSALCISQWMSGTGRKRAAKRKAQAOTIRLRGF-----ILRHSPCENA----	787
Db	728	KRSQVIAWYHRYAQOKRYOQIKSSALVJOSYIRGKAKKILRELKHOKRCEAAVTTTA	787
Qy	788	-----FLLDHVASLNLRLQPR-NVJLDSMTPRPA	820
Db	788	ATWBGTOVREYRKFFRANAGKJYEFTLORJYOKYLEKKNNKPSLSPIDKNMSRPLY	847
Qy	821	LREASELRELLOMKNMV---KYCRSISPWKOOLOKAVASEIFGKGDNPQSVPRLF	877
Db	848	FLDSTH--KELKJFLHMRCKKRYDQOTDQOKLJYEKLEASLFLDKKALYPSSYGOF	905
Qy	878	ISTRJGTEISP---RVLQSLSEPIQYAVPVYKIDKRGYKRRPOLLTPESAUIVE-	932
Db	906	GCAYL---EIKNPKRYKKLKAIEKJITIAEYVKNINRANGSTSRIFLTLNNNLLADQ	962
Qy	933	-DAKYKORIDYANLTGJVSLSLDSLEVLHVQR-EDNKKQGDVJLDSHVIEETLR---T	987
Db	963	KSGQJLKSVPJLADVTKYVSMSSQNDGFAVHLKRGESMAKSGOFLESSHILIEMAKRLYT	1022
Qy	988	ALSADRV-NINI-----NOSITPAGPGRGDGIIDTPSSSELLTKAKNGHLAV	1036
Db	1023	TLSPQKQKLNIEIDELVLVQFRODKVCVKTIQGNQKNVSP-----TCKRKNRLLLE	1074
Qy	1037	VA 1038	
Db	1075	VA 1076	

## RESULT 9

myosin heavy chain I, brush border - bovine  
 N:Alternate names: myosin I heavy chain-like protein, MIMC  
 N:contains: myosin ATPase (EC 3.6.4.1)  
 C:Species: Bos primigenius taurus (catle)  
 C:date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_revision 19-Apr-2002  
 C:accession: A29483; JX0208  
 R:Hoshimaru, M.; Nakashishi, S.  
 J. Biol. Chem. 262, 14625-14632, 1987  
 A:title: Identification of a new type of mammalian myosin heavy chain by molecular cl  
 A:reference number: A29483; MUID:88033016; PMID:3667594  
 A:accession: A29483  
 A:molecule type: mRNA  
 A:residues: 1-1043 <HOS>  
 A:cross-references: GB:J02819; NID:g163405; PIDN:AAA30658.1; PID:g163406  
 R:Kawakami, H.; Moriyoshi, K.; Utsunmi, T.; Nakashishi, S.  
 J. Biochem. 111, 302-309, 1992  
 A:title: Structural organization and expression of the gene for bovine myosin I heavy  
 A:reference number: JX0208; MUID:92268028; PMID:1587791  
 A:accession: JX0208  
 A:molecule type: DNA  
 A:residues: 1-789 <KAM>  
 A:note: nucleotide sequence is not complete  
 C:genetics:  
 A:introns: 38/3; 77/2; 109/1; 144/1; 159/3; 181/1; 214/1; 248/3; 298/1; 337/3; 366/3;  
 C:superfamily: brush border myosin heavy chain I; myosin motor domain homology

C:Keywords: actin binding; hydrolase; intestine; nucleotide binding; P-loop  
 F:11-681/Domain: myosin motor domain homology <MMOT>  
 F:101-108/Region: nucleotide-binding motif A (P-loop)  
 F:572-593/Region: actin binding #status predicted  
 F:723-1043/Domain: carboxyl-terminal <CTD>  
 F:107/Binding site: ATP (Lys) #status predicted

Query Match 39.0%; Score 2106.5; DB 1; Length 1043;  
 Best Local Similarity 44.0%; Pred. No. 1e-126;  
 Matches 455; Conservative 185; Mismatches 330; Indels 65; Gaps 21;

```

QY 27 VGVQDFVLLNFTSEAFIENLRFRREMLITYTIGPVLSVNPYDLOIYSQHMER 86
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8 VGVEDVLLLEPLEQ-SLRNLQLRREKEITYTIGNVLSVNPYQDLPIYDLEFAK 66

QY 87 GVSFEVPHLPAVDVYRALRTERDOAVMISGESGAKTEATRRLOFYAETCPAPE 146
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 67 DTFPELKHITVALANMAVQLRDRDQCILITGESGAKTEASKLVSVAAYGCKEE 126

QY 147 RCGAVRDLQSNPVLEAFGNKTLRNDSSRFSGKYMDVQDFKGPVGGHLLSYLLEKS 206
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 127 QVNSVKEQLQSNPVLEAFGNKTLRNDSSRFSGKYMDIEPFKGPGLGVTNYLLEKS 186

QY 207 RVVQHNGERNHFVYQLLEGGEETLRGLGERNPQSLYLKGGCAVSSINDKSDMK 266
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 187 RVVQKLEGERNHFIFQLLAGADQLKALKLERDGYAYL-NPDTSRVDGDDDAKPK 245

QY 267 VNRKALSVIDFDEVEDLSTIYASVLHGNLHFADEDSN---AQVTTENQKLXLT 322
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 246 VQASAMTVGFSDERQVLEVAALVAKGNVELINEPANGVPSGIDGKGVQIGEL 305

QY 323 LGVEGTTLEALTHRKTIKAGEELLSPNLDEQAAYARDALAKAVSRFTWLRKINRS 382
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 306 VGLNSVELERALSPTMETAKKVVTLNVIAQVYARDALAKAVSRFTWLRKINRS 365

QY 383 ASKDAEPMWRSTTVLGLLDIYGFVQHNSEFOFCINCNKLOQLFTELTKSQDEEY 442
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 366 KVGTCG---KRWKVGVDIYGFTELLEDNSFEQFVINCNEKLOQVLEMTLKEDEEY 420

QY 443 EAGGIMEPVQYFNNKIICDLVEEKFGIISLDECLRGEGATDLPLEKLEDTYKPPH 502
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 421 KRGGIMVAYVEFDNGICITNLIEHNRGLIAMDDECLRPVYSDSTFLAKLQFLSKHS 480

QY 503 HFLTHRLADQKTRK---SLDRGEFRLHYAGEVYTSVTEGLDKNNDLFRNLKETKCSM 559
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 481 HV-ESKVTQNAORQYHSHGSLGSCFICHYAGKYTVNVSFIDKNNDLFRDLDSQAMKAR 539

QY 560 NPIMAOCFPKSE--LSDKRPETVATQFKMSLQVLEILRSKPAYIRICIKPNDAKOPGR 617
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 540 HPLRLSLRPFEGPQKQSLKRPYAGQAFKSSVYTTLMKNLYSKNPYIRICIKPNEHQORGH 599

QY 618 FDEVLRLHGVKTLGLMENLRVRAGAFYRRKYEAFLQYRKSLCPEYPMWAGROGVAV 677
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 600 FSEFELSVQAQYVGLLENRRRAGATYRQAVGSLFERYRLLSRSTWPRNGGDDGVVER 659

QY 678 LVYHLYGKPEEYKMGRTKIFIRPKTLFATEDSELEVRSOLATKIOAARGFHWROKFLR 737
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 660 VIGELSMSEBELAFGKTKIFIRSPKTLFYLEBGRRLRLDQALTLQKTRGMCGRHYQL 719

QY 738 VKRSAICIOSMWRGTLGRKKAARWAAOTIRRLIGFLIRHSR-----CPENAF 788
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 720 MRSQIVISWFGNMOKHYRKMKASALLIOAFVYGMKARKNYRKYFRSGAALLISNFI 779

QY 789 PLDHYASFLNLRQLRP-NVLDTSMPTR-----PALREASELLRELCKNMWYKGR 842
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 780 YKSMVO-KFLIGLKNDLPSPSTIDKWPSPAPKYENTANHELRHQ-----WK-CK 830

QY 843 ----SISPEMKOOLQKAAASEIFKGDNDYQSVRLFTISRLGTEESPRVLOSLSGE 898
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 831 KRRODQSPKQVLETRKCLASLFFKCKKASYQSVPIPFHGDYIGQR-NPKIQIKLKGGE 889

QY 899 --PIQYAVPVVYDKRGYPRPRQLLTPSAVI--VEDAKYKORIDVANTLGISVSSLS 954
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 890 EGFILMAETVYKVNKGNAKTSRILLITLGHVITIDMKNPQAKTVPLNSLAGVSTSK 949

QY 955 DSLFVLYHOREDN-KQGDVVLQSDHVIETLRK---TALSADRVN-----NINIQ 1001
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 950 DGLFSLHLEISVSGSGFELLVSEHVELLTKICATLIDATQMLPVTVEEFSVKFME 1009

QY 1002 GSIT--FAGPGGRDG 1014
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1010 GSITVYKVIQPGGCG 1024

```

## RESULT 10

C45439  
 myosin-I, myr 1b (alternatively spliced) - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 02-Feb-2001  
 C:Accession: C45439  
 R:Ruppert, C., Kroschewski, R., Bahler, M.  
 J. Cell Biol. 120, 1393-1403, 1993  
 A:Title: Identification, characterization and cloning of myr 1, a mammalian myosin-I.  
 A:Reference number: A45439; PMID:8449985  
 A:Accession: C45439  
 A:Status: preliminary  
 A:Molecule type: mRNA; protein  
 A:Residues: 1-1107 <RUP>  
 A:Note: sequence extracted from NCBI backbone (NCBIN:128956, NCBI:128984)  
 C:Superfamily: brush border myosin heavy chain I; myosin motor domain homology  
 C:Keywords: nucleotide binding; P-loop  
 F:18-688/Domain: myosin motor domain homology <MMOT>  
 F:108-115/Region: nucleotide-binding motif A (P-loop)

Query Match 38.8%; Score 2094.5; DB 2; Length 1107;  
 Best Local Similarity 41.7%; Pred. No. 6.6e-126;  
 Matches 463; Conservative 183; Mismatches 346; Indels 119; Gaps 19;

```

QY 27 VGVQDFVLLNFTSEAFIENLRFRREMLITYTIGPVLSVNPYDLOIYSQHMER 86
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 15 IGVGDVLLLEPL-NEETFDNLKRRDHEIITYIGSVISVNPYSLPIYSPEKVEDYR 73

QY 87 GVSFEVPHLPAVDVYRALRTERDOAVMISGESGAKTEATRRLOFYAETCPAPE 146
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 74 NRPFEPLPHIFLSDPANKSLRDQKDCILITGESGAKTEASKLVSVAAYGCKGA 133

QY 147 RCGAVRDLQSNPVLEAFGNKTLRNDSSRFSGKYMDVQDFKGPVGGHLLSYLLEKS 206
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 134 EVNQVKEQLQSNPVLEAFGNKTLRNDSSRFSGKYMDIEPFKGPGLGVTNYLLEKS 193

QY 207 RVVQHNGERNHFVYQLLEGGEETLRGLGERNPQSLYLKGGCAVSSINDKSDMK 266
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 194 RVVQKLEGERNHFIFQLLAGADQLKALKLERDGYAYL-SLDSAKVNGVDDAANFR 252

QY 267 VNRKALSVIDFDEVEDLSTIYASVLHGNLHFADEDSN---AQVTTENQKLXLT 322
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 253 YVANAQIYGFSDPEESVLEVAALVAKGNVELINEPANGVPSGIDGKGVQIGEL 312

QY 323 LGVEGTTLEALTHRKTIKAGEELLSPNLDEQAAYARDALAKAVSRFTWLRKINRS 382
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 313 TSTDQVYLEAEPFRVYKAEKQEVSTTLNVAQYVARDALAKLHLSRLSMVLRINESI 372

QY 383 ASKDAEPMWRSTTVLGLLDIYGFVQHNSEFOFCINCNKLOQLFTELTKSQDEEY 442
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 373 KAQTKVRKK-----VMGVLDIYGFELFEDNSFEQFVINCNEKLOQVLEMTLKEDEEY 427

QY 443 EAGGIMEPVQYFNNKIICDLVEEKFGIISLDECLRGEGATDLPLEKLEDTYKPPH 502
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 428 IRREDIEMTHIDFNNKIICDLLENNTNGILAMDDECLRGVYTDLPLEKLNQVCATIQ 487

QY 503 HFLTH--KLADQKTRKSLDRGEFRLHYAGEVYTSVTEGLDKNNDLFRNLKETKCSM 560
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 488 HPESSKSKSRFLNDPTTLRHSCTRIQHYAKVLYQVEGYDKNNDLIRKDLDSAMKAGH 547

QY 561 PIMAOCFDSKSELS--DKKRPETVATQFKMSLQVLEILRSKPAYIRICIKPNDAKOPGR 618
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 548 ALIKSLFPEGNPAPKVNLRKRPPTAGSOPKASVATLMKNLQTKNPNTIRCIRKPNDKKAHIF 607  
 Qy 619 DEVLIRHQQVYGLGEMENLRVRAGFAVRRKYEAFLORYKSLCEPTWPMAGRPDGVAVL 678  
 Db 608 SESLVCHQIRYGLLENNVRRAGTAFOAYEPCLEERYMCLKQVPMHKGARSGVEVL 667  
 Qy 679 VRIILGKPEEYKMGRTKIFIRPEKTLFATEDSLERQSLATKIOAMGFHMROKFLRV 738  
 Db 668 FNELEIPVEEYSGRSKIFIRNPRTLFOLEDLRKQLEDLATLQIKYGMWCRTHFLM 727  
 Qy 739 KSAITIOSMWRGTGRRKAARKNAQITRLIRGF-----ILRHSPCPENA---- 787  
 Db 728 KSOVVIAWRYAOQKRYQIKSSALVIOSYIRGMKARKILRELKQKCKEAAATITIA 787  
 Qy 788 -----FELDH 792  
 Db 788 AVMHGTOARRELRLKLEERKRNHAAVIMAVYGLKVRRETKKFFRANAGKKEEFTLOR 847  
 Qy 793 VASFLNLRLRQLPR-NVLDTSWPTPPALREASSELLRELCKMNMV--KYCRSISPFW 848  
 Db 848 IVQKTLLEKKNMPSLSPIDKNMPSRPLYLFLDSTH--KELKRIFILMCKKYRDOFTDQ 905  
 Qy 849 KOOLQAKVASEIFKGGKDNYPQSVPRLFISTRLGTEELSP----RVLSQSGSEPTQYAV 904  
 Db 906 KLIYEKLEASLEFMDKALYPSVGOFPQAYL--ELNNPKYKKLKDAIEEKIIIAE 962  
 Qy 905 PYVKIDRKGYRPRPQLLTPSAVIVE--DAKYKORIDVANTLGISVSLSDSLFVLHV 962  
 Db 963 VVNRINRANGKSTSRIFLTLNNLLADQKSGQIKSEVPLVDVTKVSSQNDGFFAVHL 1022  
 Qy 963 QR-EDNRKQGNVLQSDHVIETLRK---TALSADRVN-NINI-----NOGSTTFA 1007  
 Db 1023 KEGSAASKGDFLSSDHLIEMATKILKRTTILSQTKQKLINEISDFLVOFRQDVCVKFI 1082  
 Qy 1008 GCGPGHGTIDFTSGSESLITTKAKNGHLAVVA 1038  
 Db 1083 QGNQKNGSVP-----TKRKNRLEVA 1105

## RESULT 11

A45438  
 myosin I alpha chain - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 21-Sep-1993 #sequence\_rev18-Nov-1994 #text\_change 02-Feb-2001  
 C:Accession: A45438  
 R:Sherr, E.H.; Joyce, M.P.; Greene, L.A.  
 J. Cell Biol. 120, 1405-1416, 1993  
 A:Title: Mammalian myosin I alpha, I beta, and I gamma: new widely expressed genes of th  
 A:Reference number: A45438; MID:93194946; PMID:844986  
 A:Accession: A45438  
 A:Status: preliminary  
 A:Molecule type: nucleic acid  
 A:Residues: 1-1079 <SNP>  
 A:Cross-references: GB:L00923; NID:920008; PID:AAA39800.1; PID:920009  
 A:Experimental source: brain  
 A:Note: sequence extracted from NCBI backbone (NCBIN:129047, NCBI:129048)  
 C:Superfamily: brush border myosin heavy chain I; myosin motor domain homology  
 C:Keywords: nucleotide binding; P-loop  
 F:18-689/Domain: myosin motor domain homology <MMOT>  
 F:108-115/Region: nucleotide-binding motif A (P-loop)

Query Match 38.6%; Score 2081.5; DB 2; Length 1079;  
 Best Local Similarity 42.7%; Pred. No. 4,3e-125;

Matches 463; Conservative 181; Mismatches 347; Indels 93; Gaps 20;

Qy 27 VGVQDFVLENTSEAFTEENLRRENTLYTYIGPVLSVNPYRDLOYSRQHMERYR 86  
 Db 15 IGVGDVNLLEPL-NEETFTDNLKKRPDNEITYIGSVIVNYPYRSLPIYSPEKVEDYR 73  
 Qy 87 GVSFEVPRPHLFAVDVTVRALTERDQAVMISGSGAGKTEAKRLQLQEAETCPAPE 146  
 Db 74 NNNFVELSPHIFALSDPAVRSRLDQDKQCIITGESGAGKTEASKLVMSVAAVCGKA 133

Qy 147 RGAIVDRLLQSNPVLEAFNGAKTLRDNSSRFEGKYMDVOFDEKAPVGHILSTLEKS 206  
 Db 134 EVNQVKEQLQSNPVLEAFNGAKTVRDNSSRFEGKYMDIEFDFKGDPLGVISNYLLEKS 193  
 Qy 207 RVVQNNHRRFHFVYQLBEGEETLRGLGLENNPQSYLYLKGQCAVSSINDKSDMK 266  
 Db 194 RVVQNPGRNHFVYQLBEGEETLRGLGLENNPQSYLYLKGQCAVSSINDKSDMK 252  
 Qy 267 VMKRALVIDFTEDEVEDLISVAVLHNLNHPAD-----EDSAQVTTENOAKLYL 320  
 Db 253 TVRNAQIVGLDHEAFAVLEVAAYLKLGNIEKPEPSRNGIDESKIKDKIELNEKFA 312  
 Qy 321 RLVEEGTTLREALTHKKIILAKGELLSPUNLEQAAVARDALAKAVYSRTFTWLVKIR 380  
 Db 313 RPAVSK-VLVERAFSFRVTEAKREKYSTLTNAQVAYARDALAKNLSRFLSVLRINE 371  
 Qy 381 SLAKDAESPMSWSTVYGLDIYGEFVPQNSPEQPCINCYCEKIQOLFIELTLKSEOE 440  
 Db 372 SIKAQTVRRK-----VGVLDIYGEIIFEDNSPEQFITINCYCEKIQOLFIELTLKSEOE 426  
 Qy 441 EYAEGLAMPEVOYFNKILICDLVEERKGIISILDEECLRPEATDLFLEKLEEDYVP 500  
 Db 427 EYIREDEIEMTHIDYFNNAITCDLIENNTGILAMDEBCLRGVYDDEFLEKLNQVAT 486  
 Qy 501 HPHFLTH--KLADQTKRSLDRGEFRLHYAGVYTVTGFLDKNDDLFRNLKETMSS 558  
 Db 487 HQFESGMSKCSFLNDTLPSCFRIQYAGVLYQVGFDPKNDLLYRLOSOAMKA 546  
 Qy 559 MNPIMACFDSKSL--DKRPETVATQFMSLQLVELIRKEPATYICIPNAPQPC 616  
 Db 547 DHSILKSLFPEGNPAPKVNLRKRPPTAGSOPKASVATLMKNLQTKNPNTIRCIRPNDKKAH 606  
 Qy 617 RPEVLIIRHQQVYGLGEMENLRVRAGFAVRRKYEAFLORYKSLCEPTWPMAGRPDGVAVL 676  
 Db 607 SESLVCHQIRYGLLENNVRRAGTAFOAYEPCLEERYMCLKQVPMHKGARSGVEVL 666  
 Qy 677 VVNRINRANGKSTSRIFLTLNNLLADQKSGQIKSEVPLVDVTKVSSQNDGFFAVHL 1022  
 Db 727 LKRSQOVIAWRYAOQKRYQIKSSALVIOSYIRGMKARKILRELKQKCKEAAATITIA 787  
 Qy 788 -----FELDHVAVSLNLRLRQLPR-NVLDTSWPTPP 847  
 Db 788 AVMHGTOARRELRLKLEERKRNHAAVIMAVYGLKVRRETKKFFRANAGKKEEFTLOR 847  
 Qy 849 KOOLQAKVASEIFKGGKDNYPQSVPRLFISTRLGTEELSP----RVLSQSGSEPTQYAV 904  
 Db 906 KLIYEKLEASLEFMDKALYPSVGOFPQAYL--ELNNPKYKKLKDAIEEKIIIAE 962  
 Qy 905 PYVKIDRKGYRPRPQLLTPSAVIVE--DAKYKORIDVANTLGISVSLSDSLFVLHV 962  
 Db 963 VVNRINRANGKSTSRIFLTLNNLLADQKSGQIKSEVPLVDVTKVSSQNDGFFAVHL 1022  
 Qy 963 QR-EDNRKQGNVLQSDHVIETLRK---TALSADRVN-NINI-----NOGSTTFA 1007  
 Db 1023 KEGSAASKGDFLSSDHLIEMATKILKRTTILSQTKQKLINEISDFLVOFRQDVCVKFI 1082  
 Qy 1008 GCGPGHGTIDFTSGSESLITTKAKNGHLAVVA 1038  
 Db 1083 QGNQKNGSVP-----TKRKNRLEVA 1105

## RESULT 12

A45439  
 myosin I heavy chain - rat  
 C:Species: Rattus norvegicus (Norway rat)

C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 02-Feb-2001  
C:Accession: A45439, S29984  
R:Ruppert, C.; Kroschewski, R.; Bahler, M.  
J. Cell Biol. 120, 1393-1403, 1993  
A:Title: Identification, characterization and cloning of myr 1, a mammalian myosin-I.  
A:Reference number: A45439; MUID:93194945; PMID:844985  
A:Accession: A45439  
A:Status: preliminary  
A:Molecule type: mRNA, protein  
A:Residues: 1-1136 <RUP>  
A:Cross-references: GB:X68199; NID:956732; PIDN:CAA48287.1; PID:956733  
A:Note: sequence extracted from NCHI backbone (NCBIN:128956, NCBIPI:128961)  
R:Ruppert, C.  
Submitted to the EMBL Data Library, September 1992  
A:Reference number: S29984  
A:Accession: S29984  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1136 <RU2>  
A:Cross-references: EMBL:X68199; NID:956732; PIDN:CAA48287.1; PID:956733  
C:Superfamily: brush border myosin heavy chain I; myosin motor domain homology  
C:Keywords: nucleotide binding, P-loop  
F:18-688/Domain: myosin motor domain homology <MOT>  
F:108-115/Region: nucleotide-binding motif A (P-loop)

Query Match 38.5%; Score 2080; DB 2; Length 1136;  
Best Local Similarity 40.68; Pred. No. 5.9e-125;  
Matches 463; Conservative 183; Mismatches 346; Indels 148; Gaps 19;

Oy 27 VGVODFVLENNFTSEAFIENLRRFRRENIYYTYIGPVLSVNPNRYDLQIYSRQHMERYR 86  
::: ||| : |::||::| | |||| ::||| | ||| : : ||  
Db ~ 15 IGVCTVLLEP-NEEFIDNLKRRFDHNEYTYTIGSVATSNPNPYSLPIYSPEKEVEDIR 73

Oy 87 GVSFEVPPHFAVADTVYRALRTERDQAVMISGSGAGKTEATKRLLGEVAETCPAPE 146  
 :|||:|||::| |||| ::||::| | | | | | | | | | | | | | |  
Db 74 NRPNELSPHFALSBAYRSLRDQDDQCILITGESGAKTEASKLVMSTVAAVCGKA 133

QY 147 RCGAVRDLRLQSNPVLAEFGNAKTLRNDNSSRFGKYMVDQEDPFKAPVGGHLSYLEKS 2066  
:::||||| |||||:::|||||:::|||||:::|||||:::|||||  
Db 134 EVNQVEKOLLQSTPVLEAFGNAKTVRNDNSSRFGKYMDFEDPFKDPGLGYSNYLEKS 1933

QY 207 RVVHONHGERNFHFVQLLEGGEEETLRRLGLEKNPQSYLYLVKGCACAKVSSINDKSDMK 266  
 ||| | ||||| ||| | : ||| : ||  
 Db 194 RVVOPRGEREFHFVQLLSGSEELLHKIKLEDESPRNYI -SLDSAKVNCVDPAANFR 252

QY 267 VMKALSVIDTEDEVEDLSTVASVHLGNHFAEDSDN----AQVTENQLKYTRL 322

Dh 253 TVRNAMOTIVGSDPAPASVI EYVAAVIKICNTFFRDSBPWCI DESKTKDNEI KTCCEI 312

Oy 323 LGVESTTLREALTHRIKIIAKGEELLSPINLEQAAAYARDALAKAVYSRTFTWLVRKINRSL 3822  
:: | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 313 TSTDOVVI FPEESEPTEMAKKOKUVTMT NVAOVYADPDI AKNIVCSDI EGTUI UNBTNGST 3772

OY 383 ASKDAESPMSRSTTVLGLLDIVGFEVFOHNSFEOFCINVCNEKLQOLFIEITLKSEQEY 4422  
::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
Dh 373 KNOGVUBYK- - - - WNCU DYCDECFEDNCSFECDTICWCEVTCOITTTIETIKCCECCY 4423

```
OY      443 EAEGLAEPVQYNFKKICDLVEEKFGKIISLDEECRLRGEAATD/LTFLXKL EDTVPKPHP 502Z  
| | | : ||||| : | || : ||||||||| || ||||| |  
DB      438 TROEDFUTUETUUWIIITCOT FURNUWTU LAY DDDO DOOUUEEEM ETI NUNUENUS 467
```

QY 503 HFLTH--KLADQKTRKSLDRGEFRLHYAGEVWYVSYTGFLDKNNDLFRNLKETWCSMN 560

```
QY      561 PMAQCDFKSELS--DKRPETVATQFKMSLQVEILRSKEPAYIRCIKPNDAKQGRF 618
        :: | : : ||| : ||| : |::| | ||||| |
E10 ATCTTCGCGAGTAAATTATTTGGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
```

QY 619 DEVLIRHQVKYLGIMENLVRRRAGFAVRRKYEAFLOQRYKSLCPETWPMWACRPODQAVL 678

DD 808 SESLVCHQIRILG LLENVRVRKAGYAFKQAYEPCLERYKMLCKQIMPHWKGPARKSVEVL 86/

QY 679 VRHLGYKPEEYKMGRTKIFIRPKTLFATEDSLEVRROSLATKIQAAMRGFHWROKFLRV 738

Db 668 FNEDEIPVEEISFGRSKIFIRNPRTLQLEDLRQRQLEDLATLQIKYRGKKCRTHFLM 7277

Oy 739 KRSALCISQSWRGTLGRKKAARKKMAOTIRRLIRGF-----ILRHSPECPENA---- 7877

Db	728	KRSQVIAWYRRYAOQKRYOQIKSSALVIOSYIRGWKARKILRELKHOKCKEAAITIA	787
Ov	788	-----	787

Db	788	AYMHCTQARKERRRLDEARNKHAIAVTWAFWLGSKARELKLKEEARRKHAAVAITWAY	847
Ov	788	-----FTLDHVASFLI.NI.EROI.PR.-NVI.DTSWTPPBALR	822

```
Db      848 WLGLKVRREYRKFRRANAGKKIYEFTLÖRIVQKYLLMKMKNPSPIDKMWPSRYEFL 9077
          || :   :: :| : :| : :| : :| : |
Ov     823 FASELI PFTCMRNMTW---KVQBSTPEFKOOOKAVASFTFCGCKNDYNQDVSPRIETS 8706
```

[illegible]

```

Db 966 AYL---EINKNPYKKLDAIEKIIIAEVNKNRANGKSTSRFLLTNNLLLDQKS 102

```

```
Db 1023 GQIKSEVPLVDYTKVMSNQNGFFAVHLKEGSEASKDPFLFSSDHLIEMATKLYPTTL 108
```

```

      | : || |      :: : | | :: | :      | | | |
Db 1083 SQTAKNLIEISDEFLVQFRQDKVCVKFIQGNQKNGSVP-----TCKRRNNRLEEVA 113

```

RESULT 13  
S32404  
myosin heavy chain I, brain - mouse

C;date: 03-May-1994 #sequence\_revision 20-Feb-1995 #text\_change 02-Feb-2000  
C;Accession: S32404; S33973  
R;Koslovsky, J.S.; Qian, C.; Jiang, X.; Mercer, J.A.

A:Title: Molecular cloning of a mouse myosin I expressed in brain.  
A:Reference number: S32404; MUID:93209370; PMID:8458427  
A:Accession: S32404

A:Residues: 1-1094 <KOS>  
A:Cross-references: EMBL:X69987; NID:g312270; PIDN:CAAA9604.1; PID:g166647  
R:Mercer, J.A.

Submitted to the EMBO Data Library, January 1995  
A: Reference number: S33973  
A: Accession: S33973  
A: Molecule type: mRNA

A; Residues: 1-312, G, 15/4-102, 1/104-1094 (MEMO)  
A; Cross-references: EMBL:X69987  
C; Superfamily: brush border myosin heavy chain I; myosin motor domain homology  
C; Keywords: nucleotide binding; P-loop

Query Match 38.4% Score 2075.5 DB 2: length 1094  
F;5-6/34domain: myosin motor domain homology <MMOT>  
F;95-102/Region: nucleotide-binding motif A (P-loop)

27 VGVNDEVLINETSFAAFTEINRRRRRNTTYTGTGVTIVSNVDPYDLOIVSRHMEYR 86  
 best local similarity 41.9%; Pred. NO. 1.1e-124;  
 Matches 464; Conservative 184; Mismatches 348; Indels 112; Gaps

87 GYSEFEVDPRLFAVADMYVRALPTFRPDAAVMISGSGGACCTEATKBLIOEVAETCPADE 146

```

Db      61 NRNEYELSPHIFALSDAYRSLRDQKQCILITGESGAGTEASKLVMYVAANGCGA 120
          :|||:|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
On     147 RCGAVDRDITGSNRYTAEAGNAVYTIPNDNSCEPQYVMQVDEKGAIVUCUTISVTIKS 208

```

14/ RGGAVRDRLLQSNPVL EAFGN AKTLRNDNSSRFGKYMDVQDFDKGAPVCGHILSYLLEKS 206

Db 121 EVNVOVKBOLLQSNPVLBAFNAKTVARNDNSSRFCKYMDIEFDEKCDPLGCVISNYLLEKS 180  
 Qy 207 RVVHONHGERNFHFVYOLLGEGEETLRRLGLEENPQSYLYLVKGCAGKSSINDKSDWK 266  
 Db 181 RVVQMPGERNFHFVYOLLGEGEETLRRLGLEENPQSYLYLVKGCAGKSSINDKSDWK 266  
 Qy 267 VMKALSVIDFTEDEBDLSIYASVLIHLNIFHADEDSN---AQVTFNEDKLYTRL 322  
 Db 240 TVRNAQVGFLDHEAVLEVAVALKGNIEKPPESRVNGIDSEIKKNELKECEL 299  
 Qy 323 LGVEGTLREALTHRKIIAGCEELSPLEEQAAVARDALAKAVSKFTWLVKTRNSL 382  
 Db 300 TSLDQVLEAFSEFRTVEAKQEVSTTLNVAQVYARDAKAKMLYSRLSMVLNRINESI 359  
 Qy 383 ASKDAESPWSRTVLGLDIYGFVEVQHNSEFQFCINCNCKEQLQOLFTLELTKSOEEY 442  
 Db 360 KAOTKVRKK-----VMGVLDIYGFELFEEDNSFEQFIINCNCKEQLQOLFTLELTKSOEEY 414  
 Qy 443 EAAGTAMEPQVYNNKIIICDLVEKFKGIISIDEECLRGATDULFLEKLEDTYKPPH 502  
 Db 415 IREDIEMTHIDYFNNAIICDLLENNTNGILAMDEECLRGVTDTEFLEKLNQVCATHQ 474  
 Qy 503 HFLTH--KLADOKTRKSLRGEFRLHYAGVYTYTGFLDKNDLLEFNLKETWSSMN 560  
 Db 475 HFESRMSKCSRFNLNNTTLPHSCRFIDHYAGVLYVEGFVDKNDLLEFNLKETWSSMN 560  
 Qy 561 PMAOCFDSKELS--DKRPETVATQFKMSLQVLVEILSKRPAYIRCIKPPDAKOPGRF 618  
 Db 535 SLIKSLFEPGNPAKVLRKPTRSSQFASVATIKRNLOTPKNPTIRCIKPPDAKOPGRF 618  
 Qy 619 DEFLRHQVYKYLGLMENLVRAGAFYRRKYEAFIQRKSLCPEPMWAGRPDQAVL 678  
 Db 595 NSLVLCHQRTYGLLENVAVRAGAFYRRKYEAFIQRKSLCPEPMWAGRPDQAVL 678  
 Qy 679 VRIHLYKPEEYKMGKTKIFIRPKTLFATEDSLEVRQSLATKIOAANRGMKRLRY 738  
 Db 655 FNELEIPVEEHSEFGSKIFIRNPRFLFOLEDKRLQLEDLTLQIKIRGMKCRTHFLM 714  
 Qy 739 KRSALCISQSMWRTGLGRKKAARKMAAOTIRRLIRG-----ILRHSPPCENA--- 787  
 Db 715 KRSQVVIAMVRRYQOKRYQOIKSSALVIOYIRGMKARKILRELKQKCKEATJIA 774  
 Qy 788 -----FELDH 792  
 Db 775 AYMHGTQARRELKRLKEARNDVAVAIWAIVGLKAVREKPPRANAGKIIYFTLOR 834  
 Qy 793 VRASLTLNLRQOLPR--NVLDTSWPTPPALREASELLRELCKMNMWV---KYCRSISPEW 848  
 Db 835 IVOKYLLKMKMKMPSLSPIDKNMPSRPLYFLDSTH--KELKRIFHLMKCKKYRQOFTDQ 892  
 Qy 849 KOOLQOKAVASEIFKGGKNDYPOVRPLFTIRLCTEISF---RVLOSIGSEPIQAV 904  
 Db 893 KLITEKLESELKDKKALPSSVGPFOGAVL--EINKNPYKRLKDAIEKIIIAE 949  
 Qy 905 PVVYDRKGYPRPROLITPSAVIVE--DAKYKORIDANLNGISVSSLSLDFLHV 962  
 Db 950 VVKNINRANGKSTRIFLTNNLLDQKSGQIKSEVPLVDYTKSMSSQNGEFTVHL 1009  
 Qy 963 QR-EDNOKGQVYLQSDHVIETLT---TALSADRVN--NINOGSITTFAGCGRGI-1 1016  
 Db 1010 KEGSEAAKSGDFLEFSSDHLIEMATKLYRTLSTQKOKLNEI---SDFLVORQDKVCV 1066  
 Qy 1017 DFTSGSELLTKKAKNGHLAVVAAPLNSR 1044  
 Db 1067 KFTQGNQ-----KNGSVR--TCRKRKNR 1087

RESULT 14  
 A33620  
 myosin heavy chain I, brush border - chicken (fragment)  
 N:Contains: myosin ATPase (EC 3.6.4.1)  
 C:Species: Gallus gallus (chicken)  
 C>Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 19-Apr-2002

C:Accession: A33620  
 R:Garcia, A.; Coudrier, E.; Carboni, J.; Anderson, J.; Vandekerckhove, J.; Mosseker, M.  
 J. Cell Biol. 109, 2895-2903, 1989  
 A:Title: Partial deduced sequence of the 110-kD-calmodulin complex of the avian Intes  
 A:Reference number: A33620; MUID:90078325; PMID:2687288  
 A:Accession: A33620  
 A:Molecule type: mRNA  
 A:Residues: 1-1000 <GAR>  
 A:Cross-references: GB:X58479; NID:963075; 'PIDN:CAA41388.1; PID:963076  
 C:Superfamily: brush border myosin heavy chain I; myosin motor domain homology  
 C:Keywords: actin binding; hydrolase; intestine; nucleotide binding; P-loop  
 F:1-69/Domain: myosin motor domain homology <MMO>  
 F:59-66/Region: nucleotide-binding motif A (P-loop)  
 F:530-551/Region: actin binding #status predicted  
 F:681-1000/Domain: carboxyl-terminal <CTD>  
 F:65/Binding site: ATP (lys) #status predicted

Query Match 35.3%; Score 1904; DB 1; Length 1000;

Best Local Similarity 41.6%; Pred. No. 9.2e-114;

Matches 415; Conservative 191; Mismatches 358; Indels 34; Gaps 16;

Qy 64 VLVSVNPPYRDIQIYSRQIMERYRGVFEYVPPHLEFAVDIYVRLATREBRDQAMISGES 123  
 Db 2 VVISVNPYKPLPYITPEVVEEYHNCNFFAVKPHIYALDADYRSLRDRDQCLTIGES 61  
 Qy 124 GAGKTEATKRLLOFYAETCPAPERGAVERDRLOSNPVLBAFNAKTVARNDNSSRFCKYM 183  
 Db 62 GAGKTEASKLVMSVYAAVSSKGEVDYKRBQLOSNPVLBAFNAKTVARNDNSSRFCKYM 121  
 Qy 184 DVQDFGKAPRGSHILSYLLEKSRVYVHONNGERFHFYQLLBGGEETLRRLGLEENPQ 243  
 Db 122 DVQDFGKAPRGSHILSYLLEKSRVYVHONNGERFHFYQLLBGGEETLRRLGLEENPQ 243  
 Qy 244 SYLVLYKQGCAYSSINDKSDWKVYRKALSVIDFTEDEBDLSIYASVLIHLN---H 299  
 Db 182 HGYTL-NHEKSVTLREGMDDANFRAMQDAMALIGPAPEVATLLEVAVALKGNVLS 240  
 Qy 300 FAATEDSNAOYTTENOAKYLIRLLGVTGLRLRLHAKIIAGCEELSPLENEQAAYAR 359  
 Db 241 FOASGMASSTIAERPLEDETSLGLDPSLEALQSRVYKVDSEVLTALSVQGYGR 300  
 Qy 360 DALAKAVYSRTFTWLVKIRNSLASKDAESPWSRTVLGLDIYGFVEVQHNSEFQFCI 419  
 Db 301 DALAKNIYSRLFDMLVRINTSIQVKG-----KQKVMQVLDIYGFELFQDNGFEQFI 355  
 Qy 420 NVCNEKIQQLFELITLKSDEBEYEAAGIAMEPVQYVNNKIIICDLVEKFKGIISIDEE 479  
 Db 356 NVCNEKIQQLFELITLKSDEBEYEAAGIAMEPVQYVNNKIIICDLVEKFKGIISIDEE 479  
 Qy 480 LRGEATDLEFLEKLEDTYKPPHFLHKLADOK--TRKSLDRGEFRLHYAGVYTYST 537  
 Db 416 LRGTAVNEDEFTIKLQOIRASHKRIESELKLNKHTDVSPLRGCFRIHNYAGKYVNT 475  
 Qy 538 GFLDKNNDLFRNLKJKTQCSMNPIMAOCEFDK--ELSDKRPETVATQFKMSLQVLVE 595  
 Db 476 GFLEKNNDLFRDLQAMMAARHTLLRSLEFPGDPQRPDLKPLPTTGQFKAQSVATLKN 535  
 Qy 596 LRSKEPARYIRCIKPPDAKOPGRDEVLIRHOKYKYLGLMENLVRAGAFYRRKYEAFI 655  
 Db 536 LRSKEPARYIRCIKPPDAKOPGRDEVLIRHOKYKYLGLMENLVRAGAFYRRKYEAFI 655  
 Qy 656 YKSLCEWPMAGRPDQAVLVNHLGKYPPEYKMGKTKIFIRPKTLFATEDSLEVR 715  
 Db 596 YKSLCEWPMAGRPDQAVLVNHLGKYPPEYKMGKTKIFIRPKTLFATEDSLEVR 715  
 Qy 716 OSIATKIOAAMGEFHMOKFLVVKSAICIOSMWRGTLGRKKAARKMAAOTIRRLIRG 775  
 Db 656 ABLATLIQIMFRCGCMCRKRYOLMRKQILISAMFGHQRNRYQMKRSVLLDQAYARG 715  
 Qy 776 ILR-----HSPRCPE--NAFELDHVRSFLLNLRQL--PRNVLDTSWPTPP--PALREAS 825  
 Db 716 KTRMYRRTYFNSDACTRLSNFTYRRMYQYLMGLKNLPRMAVLDRTWPPAPYKFLSDAN 775

QY 826 ELRLRLCKNNWVKYCRSISPENKQOOLQKAVASELFGKKNNYQSVPRLEFISTRIG-T 884  
 Db 776 QELKSIFYRMKCKKREQLTPQORAMLOAKLMPRLQFLDKKALVLOSLQOPFCEYLGLT 835  
 QY 885 EELSPRLVQSLGSEPIQYAVPVKVKDRKGKPRPROLLITPSAVVIVED--AKVQRIDY 942  
 Db 836 QNRKQKQKQAKKDLVMAEAVQKVRANKGTVPRLRLITTEHLVLADKKAQPKMVLSTL 895  
 QY 943 ANLTGIVSVSLSDSLFVLHVOREDNK-QKGVVLOSDHVIETLRK--TALSAD-RVNNI 997  
 Db 896 CDIQGASVSRSDGLALHLKETSITAGCKGDLIVSPHIELVTLHQTLMATQALPL 955  
 QY 998 NI-NOGSTTFRAGGPGCDGIDFT-SCSELLTKAKNGH 1033  
 Db 956 SIADQSTFRFGDVAVTVESAKGCGVPYCKKRGSH 993

## RESULT 15

A53933  
 myosin I myr 4 - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 07-Oct-1994 #sequence\_revision 07-Oct-1994 #text\_change 02-Feb-2001  
 C:Accession: A53933  
 R:Baehler, M.; Kroschewski, R.; Stoeckli, H.E.; Behrmann, T.  
 J. Cell Biol. 126, 375-389, 1994  
 A:Title: Rat myr 4 defines a novel subclass of myosin I: identification, distribution, and function  
 A:Reference number: A53933; MUID:94308268; PMID:8034741  
 A:Accession: A53933  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1006 <BAE>  
 A:Cross-references: GB:X11997; NID:9516395; PID:9516396  
 C:Superfamily: brush border myosin heavy chain I; myosin motor domain homology  
 C:Keywords: nucleotide binding; P-loop  
 F:12-682/Domain: myosin motor domain homology <MMOT>  
 F:102-109/Region: nucleotide-binding motif A (P-loop)

Query Match 34.5%; Score 1861; DB 2; Length 1006;  
 Best Local Similarity 41.7%; Pred. No. 5.3e-111;  
 Matches 426; Conservative 158; Mismatches 301; Indels 136; Gaps 26;

QY 28 GVQDVVLENFSEAFENLRARRRRLITYIGPVLSVNPYRDLYISQHMERYNG 87  
 Db 10 GRADVLMDT-VSMPEFMANLRLREFKRIYFIEGVVSNVPYVNLVYGRDTIYQYKG 68  
 QY 88 VSFYVPPHLPVADTVVRLRTERDQAVMISGESGAGTATKRLQFVATCPAPER 147  
 Db 69 RELYERPHLPFLIDAAIKAMRRSKDTCIMISGESGAGTASKTIIMOYIAITNPQR 128  
 QY 148 G--GAVRDLLOSQNVLEAFNAKTLRNDSSRFKYMVDQDFKGAQVGGHLSYLEK 205  
 Db 129 AEIEHVKNNMLKSNCLVLEAFNAKTLRNDSSRFKYMINDFKDPIGHIINNYLEK 188  
 QY 206 SPVAVNQNGERFHFYQQLLEGEEETLRGLERNPQSYLYLVKGQCAKVSINDKSDW 265  
 Db 189 SKVIVQQPEERSFHSFYQLQGSSEOMLHSLQKSSLSYNYIRVGAQLK-SSINDAAEF 247  
 QY 266 KYMRKALSVIDTEDEVEDLISVASVLHGNIFPADEDS---NAQVTENOLKYLTR 321  
 Db 248 KYVAADAMKVIYIGKPREIQTYYKILAILHLGNLKFIVDGDPLIENGKY-----VSVAE 302  
 QY 322 LIGVEGTLREALTHRKIIIAKGEELLSPNLNQ-AAYARDALAKAVYSRTFMVLRKINR 380  
 Db 303 LLSTRADWYERKALLR-IVATGRDIIIDKQHTQDEASYGDAFAKAIYERLFCMIYTRIND 361  
 QY 381 SLASDAESPSRSTTVGLLDIYGFVQVHNSFEQFCINCYENKLOQLFIETLTKSEOE 440  
 Db 362 ILEVNYDTYVHAKNTYGVLDIYGFELFDNNSFEQFCINCYENKLOQLFIQVLVQEOE 421  
 QY 441 EYEAGIAMEPQYFNKKIICDVEKFKGIISLDEBCLRGESATDLFLEKLEDTVKP 500  
 Db 422 EYQREGIIPKHIIDYFNNOITVDLVEQOHKGIITAILDDACMANGKYTDGMFLALNSKLG 481

QY 501 HPHFLTHKLADOKTRKSLDRGEPRLLHYAGEVTVSYTGFLDKNDLLEFNLEKTMGSSAN 560  
 Db 482 HGHFSSRKTCASDKILEFDR-DEFRIHYAGDVVYSVIGFIDKKKDTLLFODFKRLMYNSSN 540  
 QY 561 PTMAQCFDSELS---DKRREPVAQOFKMSLLQLEIRSKREPARIQCKPDAQOPGR 617  
 Db 541 PYLKNMPEPKLSTIEVTRPLATLFKNSMALVDNLASKPEYVVKCPKPDKRSPOI 600  
 QY 618 FDEVLIRHOKVILGLMENLRVRAGFAVRRKYEAFLQRYKSLCPETPMWAGRPQGVAV 677  
 Db 601 FDEDERRHQVEYLGLEENVRVAGFAFQYTEKFLHRYKMISEFTWPN-HDLPSQKEAV 659  
 QY 678 --LVRLHYKPEEYKMGRTKIFIRPKTLFATEDSLVBRQSLATKIQAAWGRGFWRKQF 735  
 Db 660 KKLIERCGQ-DVAVAGTKKIFIRTPRTLTLE---ELRAQMLV----- 699  
 QY 736 LRVKRSALCIQSMWRGTLGRKKAARKNAQTRILRLRGLRHSRCPENAFPDHVA 795  
 Db 700 ---RVVLFLOKVMRGTLABMR-YKFTKALVTIRYRRYKVK----- 737  
 QY 796 SFPLNLRQL--PRNVL-----TSWPPPALREASELRLQCMNNWVKYCRSISPEWK 849  
 Db 738 SYIHEVARRFHGVKNNRQDGKHHKWTTPKVLRFEEALQSLFNNRRASQLIKTIPASDL 797  
 QY 850 QOLQOKAVASELFGKKNDYPOSVPRLITSTRIGTEISPRVLQSLGSEPIQYAVPV-- 907  
 Db 798 PQVRAKVAEMMLKQGRADL--GLQRAWGNVLAASKPDTP--QTSQT-----FVYVANE 847  
 QY 908 -----KY-----DRKGKPRPROLLITPSAVVIVEDAKY 936  
 Db 848 LRRKDYMNVLFSCHAVKVNRRSKVEDRAIIVTDRLHYKMDP-----TQYKY 895  
 QY 937 KORIDYANLTGISVSLSDSLFVLHVOREDNKQKGVVLOSDHVIETLRKTAASDRVNN 996  
 Db 896 KMTIPLYNLTGLSVSGKQQLVYFHT--KDNK-----DLIVCLFSQPTHESRIGE 944  
 QY 997 I 997  
 Db 945 L 945

Search completed: December 9, 2002, 16:47:47  
 Job time : 49 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 9, 2002, 16:37:56 ; Search time 25 Seconds  
(without alignments)  
1732.051 Million cell updates/sec

Title: US-09-893-371-1  
Perfect score: 5398  
Sequence: 1 MRYRASALGSDGVVTWESA.....LITKAKNGHLAVVAPRLNSR 1044

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt.40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	5320	98.6	1028	MYIC_MOUSE	O9wt17 mus musculus
2	5143	95.3	1028	MYIC_HUMAN	O00159 homo sapien
3	2626	48.6	1026	MYIB_DROME	O23979 drosophila
4	2106.5	39.0	1043	MYHL_BOVIN	P10568 bos taurus
5	2090.5	38.7	1107	MYIA_MOUSE	P46735 mus musculus
6	2080	38.5	1136	MYIA_MOUSE	O05096 rattus norv
7	2064.5	38.2	1043	MYHL_HUMAN	O9dbcs homo sapien
8	2005.5	37.2	1045	MYHL_CHICK	P47807 gallus gall
9	1987	36.8	909	MYHL_MOUSE	O88329 mus musculus
10	1932.5	35.8	842	MYHL_RAT	O62774 rattus norv
11	1800.5	33.4	1011	MYIA_DROME	O23978 drosophila
12	1739	32.2	1003	MYSE_DICDI	P70248 mus musculus
13	1726.5	32.0	1099	MYSE_DICDI	P34092 drosophila
14	1692.5	31.4	1111	MYSB_DICDI	P34109 drosophila
15	1601	29.7	1113	MYSD_DICDI	P22467 drosophila
16	1588.5	29.4	994	MYSA_DICDI	P19766 drosophila
17	1574.5	29.2	1147	MYSB_DICDI	P10569 acanthameob
18	1523.5	28.2	1168	MYSC_ACACA	P10569 acanthameob
19	1501	27.8	1219	MYZ9_YEAST	O04439 saccharomyc
20	1453	26.9	1281	MYSC_DICDI	P45329 drosophila
21	1427.5	26.4	1273	MYSC_YEAST	P36006 saccharomyc
22	1354	25.1	2215	MY7A_HUMAN	O14402 homo sapien
23	1325	24.5	2052	MY10_BOVIN	P79114 bos taurus
24	1324.5	24.5	2052	MY7A_MOUSE	P97479 mus musculus
25	1284.5	23.8	2058	MY10_HUMAN	O9nd67 homo sapien
26	1279	23.7	2116	MYS2_DICDI	P08799 drosophila
27	1263	23.4	1742	MYSC_HUMAN	O9nqx4 homo sapien
28	1259	23.3	3511	MY15_MOUSE	O9qz44 mus musculus
29	1243.5	23.0	3530	MY15_HUMAN	O9nkn7 homo sapien
30	1233.5	22.9	1509	MYSN_ACACA	P05659 acanthameob
31	1225.5	22.7	1853	MY5A_MOUSE	O99104 mus musculus
32	1224.5	22.7	1828	MY5A_RAT	O9gyf3 rattus norv
33	1224.5	22.7	2245	MY5J_DICDI	P54697 drosophila

34	1217.5	22.6	1938	MY5A_MOUSE	P24733 aequipecten
35	1217	22.5	1855	MY5A_HUMAN	O9y411 homo sapien
36	1216.5	22.5	1846	MY5B_RAT	P70569 rattus norv
37	1206	22.3	1829	MY5A_CHICK	O02440 gallus gall
38	1206	22.3	1978	MYHB_CHICK	P10587 gallus gall
39	1202	22.3	1935	MYH7_PIG	P79293 sus scrofa
40	1198	22.2	1935	MYH7_RAT	P02564 rattus norv
41	1192	22.1	1935	MYH7_HUMAN	P35749 homo sapien
42	1192	22.1	2017	MYSN_DROME	O99323 drosophila
43	1191	22.1	1939	MYH6_HUMAN	P13533 homo sapien
44	1190	22.0	1972	MYHB_RABIT	P35748 oryctolagus
45	1187.5	22.0	1976	MYHA_BOVIN	O27991 bos taurus

## ALIGNMENTS

RESULT 1  
ID MYIC\_MOUSE STANDARD: PRT: 1028 AA.  
AC O9wt17: 008834: 008571: O9QW54:  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Myosin Ic (Myosin I beta) (Mmbp).  
GN MYOIC.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OC NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS A AND B).  
RC STRAIN=DNA/2J; TISSUE=Brain;  
RX MEDLINE=97325785; PubMed=9182797;  
RA Hamilton B.A., Smith D.J., Mueller K.L., Kerrebrock A.W.,  
RA Bronson R.T., van Berkel V., Daly M.J., Kruglyak L., Reeve M.P.,  
RA Nemhauser J.L., Hawkins T.L., Rubin E.M., Lander E.S.;  
RT "The vibrator mutation causes neurodegeneration via reduced expression  
RT of Pimp alpha: positional complementation cloning and extragenic  
RT suppression.";  
RT Neuron 18:711-722(1997).  
RL [2]  
RL SEQUENCE FROM N.A. (ISOFORM A).  
RN STRAIN=C57BL/6J; TISSUE=Lung;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shigaawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schiml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Norone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Tovo-Oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RT Nature 409:685-690(2001).  
RL [3]  
RL SEQUENCE OF 1-807 FROM N.A. (ISOFORM B).  
RN STRAIN=BALE/c; TISSUE=cochlea;  
RX MEDLINE=97237053; PubMed=9119401;  
RA Crozet F., Amraoui A.E., Blanchard S., Lenoir M., Ripoll C., Vago P.,  
RA Hameel C., Fizames C., Levi-Acobas F., Depetris D., Mattei M.-G.,  
RA Well D., Pujol R., Petit C.;  
RT "Cloning of the genes encoding two murine and human cochlear

QY	17	MESALTAARDVGVQDVLLENTSEAFATENIRRERENLIYTYIGVLVSVNPRDLOI	76
Db	1	MESALTAARDVGVQDVLLENTSEAFATENIRRERENLIYTYIGVLVSVNPRDLOI	76
QY	77	YSQOHMERKGVSYFVEPPHLPAVADYTYALRLTERROQAMISGESGACKTATYRLQ	13
Db	61	YSQOHMERKGVSYFVEPPHLPAVADYTYALRLTERROQAMISGESGACKTATYRLQ	12
QY	137	FVAFETCPAERCGAVRDLLOSNPVLEAFGNAKTLRNDNSRFGKYMDVOFDEKAPVGC	19
Db	121	FVAFETCPAERCGAVRDLLOSNPVLEAFGNAKTLRNDNSRFGKYMDVOFDEKAPVGC	18
QY	197	HLISTYLESRVYVHONHGERNPVHFYOLBEGGEETLRIGLERNPQSTLYLVKGOCAY	25
Db	181	HLISTYLESRVYVHONHGERNPVHFYOLBEGGEETLRIGLERNPQSTLYLVKGOCAY	24
QY	257	SSINDKSMKWARKALSVIDTFEDEVEDLSIYASVYLHGINHFADDESNQVTEBOL	31
Db	241	SSINDKSMKWARKALSVIDTFEDEVEDLSIYASVYLHGINHFADDESNQVTEBOL	30
QY	317	KLUTFLLECEGTLREALTHKKIIAKEBELLSPNLBOAAYADALAKAVYSRTETWLR	37
Db	301	KLUTFLLECEGTLREALTHKKIIAKEBELLSPNLBOAAYADALAKAVYSRTETWLR	36
QY	377	KINNSLASKDAESPSMRSTYVLGLLDYGEFVEQHNSFOFCINCNKLOOLFELTLK	43
Db	361	KINNSLASKDAESPSMRSTYVLGLLDYGEFVEQHNSFOFCINCNKLOOLFELTLK	42
QY	437	SEOEFEVEREGIAMEPVYOFNNKIIICDLYBEKRGIIILDEBCLRPGEATDLTEKLED	49
Db	421	SEOEFEVEREGIAMEPVYOFNNKIIICDLYBEKRGIIILDEBCLRPGEATDLTEKLED	48
QY	497	TVKPHPHFLHLKADOKTRKSLDRGEBFLNHYAGEVTVSYNGFLOXNNDDLFLRULKETMC	55
Db	481	TVKPHPHFLHLKADOKTRKSLDRGEBFLNHYAGEVTVSYNGFLOXNNDDLFLRULKETMC	54
QY	557	SSMNPITAAOCFDSSELSDDKREPVTATOFKMSLLQVLEILRSKEBAYIRCIKPRDAKOPG	61
Db	541	SSMNPITAAOCFDSSELSDDKREPVTATOFKMSLLQVLEILRSKEBAYIRCIKPRDAKOPG	60
QY	617	RFEDEVLLRHQVYKVLGLMEMNRYVRAGFAYRRKYEAFLORYKSLCPEYPMAGRPDQVYA	67
Db	601	RFEDEVLLRHQVYKVLGLMEMNRYVRAGFAYRRKYEAFLORYKSLCPEYPMAGRPDQVYA	66
QY	677	VLYVRHLGYCRPEEYKMGRIETFRPKTFLATEDSLEVRQSLARKIOAAMRGFHMROKFL	73
Db	661	VLYVRHLGYCRPEEYKMGRIETFRPKTFLATEDSLEVRQSLARKIOAAMRGFHMROKFL	72
QY	737	RVKRSATIGCLOSMWRGTLGRKRAAKKKMAOITRLINGFLIRHSPPRCPENAFGLDHVRAS	79
Db	721	RVKRSATIGCLOSMWRGTLGRKRAAKKKMAOITRLINGFLIRHSPPRCPENAFGLDHVRAS	78
QY	797	FLINLNRLOLPRNVLDTSWPTPPALREASELIRELCMKNVMKCYRSISPENKQOOLQKA	85
Db	781	FLINLNRLOLPRNVLDTSWPTPPALREASELIRELCMKNVMKCYRSISPENKQOOLQKA	84
QY	857	VASEIFKCKKNDYPOSPVRLIFSTRLCTEELSPRVLOSLSSEPTQYAVPVYKDRKGYP	91
Db	841	VASEIFKCKKNDYPOSPVRLIFSTRLCTEELSPRVLOSLSSEPTQYAVPVYKDRKGYP	90
QY	917	RPRQLLTPSAVIVEDAKVKQRIIDYANLGTISVSSISDLSPLFVHOREDNOKQDVLQ	97
Db	901	RPRQLLTPSAVIVEDAKVKQRIIDYANLGTISVSSISDLSPLFVHOREDNOKQDVLQ	96
QY	977	SDHVEFTLTKALSDADVNNININOSITFPAGPGDGIITDTSGESELLITKANKHLAV	103
Db	961	SDHVEFTLTKALSDADVNNININOSITFPAGPGDGIITDTSGESELLITKANKHLAV	102
QY	1037	VAPRLNSR 1044	
Db	1021	VAPRLNSR 1028	

```

RESULT 2
MYIC_HUMAN STANDARD: PRT; 1028 AA.
ID 000159;
AC 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin Ic (Myosin I beta) (MIM-beta) (MIMb).
GN MYOIC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=97237053; PubMed=9119401;
RA Crozet F., Amaraoui A.E., Blanchard S., Lenoir M., Ripoll C., Vago P.,
RA Hamel C., Fizames C., Levi-Acobas F., Depetris D., Mattei M.-G.,
RA Weil D., Pujol R., Petit C.;
RT "Cloning of the genes encoding two murine and human cochlear
RT unconventional type I myosins.";
RL Genomics 40:332-341(1997).
CC -I- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE
CC ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
CC THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS
CC COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS (BY
CC SIMILARITY).
CC -I- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X98507; CAA67131.1;
DR HSSP: P08799; IMND.
DR GeneW: HGNC:7596; MYO1B.
DR GeneW: HGNC:7597; MYO1C.
DR MIM: 606538;
DR InterPro: IPR000048; IQ_region.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head; 1.
DR Pfam: PF00612; IQ; 3.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head; 1.
DR SMART: SM00015; IQ; 2.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS50096; IQ; 2.
KW Myosin; ATP-binding; Actin-binding; Calmodulin-binding; Repeat;
KW Multigene family.
FT DOMAIN 1 683 MYOSIN HEAD-LIKE.
FT DOMAIN 699 732 IQ 1.
FT DOMAIN 723 751 IQ 2.
FT NP_BIND 105 112 ATP (POTENTIAL).
SQ SEQUENCE 1028 AA; 118037 MW; 0E9C680527F85C6 CRC64;
Query Match 95.3%; Score 5143; DB 1; Length 1028;
Best Local Similarity 96.2%; Pred. No. 5; 6e-319;
Matches 989; Conservative 23; Mismatches 16; Indels 0; Gaps 0;

```

```

QY 17 MESALTARDRVGVQDFVLLNFTSEAFIENLRRENLITVYIGPVLSVNPYRDLQI 76
ID 1:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSALTARDRVGVQDFVLLNFTSEAFIENLRRENLITVYIGPVLSVNPYRDLQI 60
QY 77 YSRQHMERRGVSFYEVPPLFLFAVDVYRALRTERDDAVMISGSGAGKTATATRLQ 136
ID 1:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 YSRQHMERRGVSFYEVPPLFLFAVDVYRALRTERDDAVMISGSGAGKTATATRLQ 120

```

```

QY 137 FYAETCPAPERGAVDRDLQSNPVLAEFGNAKTLRNDSSRFKGYMDVOFDKGAUVG 196
ID 1:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 FYAETCPAPERGAVDRDLQSNPVLAEFGNAKTLRNDSSRFKGYMDVOFDKGAUVG 180
QY 197 HILSYLLEKSRVYHONHGEENFHFVQLLEGGEETLRRLGRLNRNQSXYLYLKGCAKY 256
ID 1:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 HILSYLLEKSRVYHONHGEENFHFVQLLEGGEETLRRLGRLNRNQSXYLYLKGCAKY 240
QY 257 SSINDKSDMKVMKALSVIDEFEDEVEDLLSTIVASYLHGINIFPADEDSNAQVTENOL 316
ID 1:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 241 SSINDKSDMKVMKALSVIDEFEDEVEDLLSTIVASYLHGINIFPADEDSNAQVTENOL 300
QY 317 KYTLRLGVEGTYTLREALTHRKIIAAGEELSPJNEQAAVARDALAKAVYSFTWLV 376
ID 1:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 301 KYTLRLGVEGTYTLREALTHRKIIAAGEELSPJNEQAAVARDALAKAVYSFTWLV 360
QY 377 KIRSLASKDAESPMSRTTVGLDITGFEVYQHNSEFQFCINYNCKIQLFIETLTK 436
ID 1:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 361 KIRSLASKDAESPMSRTTVGLDITGFEVYQHNSEFQFCINYNCKIQLFIETLPLK 420
QY 437 SGEETEAEGIAEPYOVFNKKIICDLVEKFKGIISILDEECLRGEATDLFPLEKED 496
ID 1:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 421 SGEETEAEGIAEPYOVFNKKIICDLVEKFKGIISILDEECLRGEATDLFPLEKED 480
QY 497 TVKPHHFLTHKLADQKTRKSLDRGEFRLHYAGEVYVYTGFLDKNNDLFRNLKETMC 556
ID 1:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 481 TVKPHHFLTHKLADQKTRKSLDRGEFRLHYAGEVYVYTGFLDKNNDLFRNLKETMC 540
QY 557 SSNAPIMAGCFDKSELSDDKRPETVATQFMSLQLVLEILRSKPAYIRICIRPKDAKOP 616
ID 1:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 541 SSNAPIMAGCFDKSELSDDKRPETVATQFMSLQLVLEILRSKPAYIRICIRPKDAKOP 600
QY 617 RPEEVLIRHGVKTLGMENTRVRAGFAVRRKYEAFLQRYKSLCEPWPWAGGPOGVA 676
ID 1:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 601 RPEEVLIRHGVKTLGMENTRVRAGFAVRRKYEAFLQRYKSLCEPWPWAGGPOGVA 660
QY 677 VLVRHLGYKPEEYKMGRTKIFIRPKTLFATEDSLEYRQSLATKIQAAWRGFHMROKFL 736
ID 1:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 661 VLVRHLGYKPEEYKMGRTKIFIRPKTLFATEDSLEYRQSLATKIQAAWRGFHMROKFL 720
QY 737 RVKRSATCIOSMWRGTLGRKAAKRWAAQTIRRLRGFLIRHSPPRPEAFLDHYRAS 796
ID 1:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 721 RVKRSATCIOSMWRGTLGRKAAKRWAAQTIRRLRGFLIRHSPPRPEAFLDHYRAS 780
QY 797 FLNLNLRQLPRNVLDTSWPPPALREASELRELCKNNWVWYCRSISPENKQOOLQKA 856
ID 1:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 781 FLNLNLRQLPRNVLDTSWPPPALREASELRELCKNNWVWYCRSISPENKQOOLQKA 840
QY 857 VASEIFRGKKNDYPOSVPRLEISTRLGTEIISPRVLOSLGSEPIQYAVPVVXYDKRGYK 916
ID 1:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 841 VASEIFRGKKNDYPOSVPRLEISTRLGTEIISPRVLOSLGSEPIQYAVPVVXYDKRGYK 900
QY 917 RPRQLLTSPSAVVIYEDAKYKORIDYANLTVGISVSSISDSLFLVHVOREDNKQGDVYLQ 976
ID 1:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 901 RPRQLLTSPSAVVIYEDAKYKORIDYANLTVGISVSSISDSLFLVHVOREDNKQGDVYLQ 960
QY 977 SDHVIFELTKTALSADRVNNININOGSITPAGRGGRGIIIDFTSGSLLLTAKKNGHLAV 1036
ID 1:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 961 SDHVIFELTKTALSADRVNNININOGSITPAGRGGRGIIIDFTSGSLLLTAKKNGHLAV 1020
QY 1037 VAPRLNSR 1044
ID 1:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1021 VAPRLNSR 1028

```

```

RESULT 3
MY1B_DROME STANDARD: PRT; 1026 AA.
ID MY1B_DROME
AC 023979; 09W0H0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Myosin IB (MIB) (brush border myosin IB) (BBMIB).

```

GN MYO61F OR CG9155.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7221;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND DEVELOPMENTAL STAGE.  
 RC TISSUE-Head;  
 RX MEDLINE=94260541; PubMed=8201616;  
 RA Morgan N.S., Skovronsky D.M., Artavanis-Tsakonas S., Mooseker M.S.;  
 RT "The molecular cloning and characterization of Drosophila melanogaster  
 myosin-1A and myosin-1B."  
 RL J. Mol. Biol. 239:347-356(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Gehlker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer V., Chapple M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Butkova D., Butcher M.R., Bouck J., Brockstein P., Brothier P.,  
 RA Burlingame K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke S., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Hochia S., Dunov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiga S., Fleischmann W.,  
 RA Fostel C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostlin D., Houston K.A., Howland T.J., Wei M.-H., Ileguam C.,  
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puft V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Soler E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Stryckas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 RN [3]  
 RP SEQUENCE OF 338-697 FROM N.A.  
 RX MEDLINE=94029958; PubMed=8216259;  
 RA Cheney C.M., Kravitz N.G., Verdsky J.W.;  
 RT "A new myosin I gene in Drosophila."  
 RL Biochem. Biophys. Res. Commun. 195:1280-1288(1993).  
 RN [4]  
 RP SEQUENCE OF 630-1026 FROM N.A.  
 RA Caggese C.;  
 RL Submitted (Jul-1997) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP FUNCTION, SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND DEVELOPMENTAL  
 STAGE.  
 RX MEDLINE=96069904; PubMed=7589814;  
 RA Morgan N.S., Heintzelman M.B., Mooseker M.S.;  
 RT "Characterization of myosin-1A and myosin-1B, two unconventional  
 myosins associated with the Drosophila brush border cytoskeleton.";

RL Dev. Biol. 172:51-71(1995).  
 CC -1- FUNCTION: INVOLVED IN DIRECTING THE MOVEMENT OF ORGANELLES ALONG  
 CC ACTIN FILAMENTS (BY SIMILARITY)  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. PROTEIN SHIFTS FROM THE  
 CC BASOLATERAL TO APICAL DOMAIN IN ENTEROCYTES AND FOLLICLE  
 CC CELLS.  
 CC -1- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED IN GASTRIC CAECAE,  
 CC MIDGUT CELLS OF THE PROVENTRICULUS, AND IN THE MID AND HINDGUT. IN  
 CC THE LARVAL AND ADULT GUT BRUSH BORDER, EXPRESSED IN THE  
 CC MICROVILLI. ALSO EXPRESSED AT HIGH LEVELS IN FOLLICLE CELLS DURING  
 CC OÖGENESIS.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSION STARTS AT 8-12 HOURS EMBRYONIC  
 CC DEVELOPMENT, CONTINUES TO INCREASE UNTIL THIRD LARVAL INSTAR.  
 CC DISAPPEARS IN PUPAE AND IS PRESENT AT A LOW LEVEL IN ADULTS.  
 CC EXPRESSION IN EMBRYOGENESIS IS CORRELATED WITH THE FORMATION OF A  
 CC BRUSH BORDER WITHIN THE ALIMENTARY CANAL.  
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 2 IO DOMAINS.  
 CC -1- CAUTION: Ref.2 sequence differs from that shown due to erroneous  
 CC gene model prediction.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>  
 CC or send an email to [license@isb-slb.ch](mailto:license@isb-slb.ch)).  
 CC -----  
 CC EMBL: 007596; AAA19591.1; -  
 CC EMBL: AE003471; AAF47477.1; ALT SEQ.  
 CC EMBL: L13070; -; NOT ANNOTATED\_CDS.  
 CC EMBL: AJ000879; CAA04367.1; -  
 CC HSSP: P08799; 1MDN.  
 CC Flybase: Fggn0010246; Myo61F.  
 CC InterPro: IPR000048; IO\_region.  
 CC InterPro: IPR001609; myosin\_head.  
 CC Pfam: PF00063; myosin\_head; 1.  
 CC Pfam: PF00612; IO; 3.  
 CC PRINTS: PR00193; MYOSINHEAVY.  
 CC PRODOM: PD00035; myosin\_head; 1.  
 CC SMART: SM00015; IO; 1.  
 CC SMART: SM00242; MYSC; 1.  
 CC PROSITE: PS50096; IO; 1.  
 CC Myosin: Arp-binding; Calmodulin-binding; Actin-binding; Repeat.  
 CC FT DOMAIN 1  
 CC FT 697 724  
 CC FT NP\_BIND 105 112 ATP (BY SIMILARITY).  
 CC FT DOMAIN 583 590 ACTIN-BINDING (BY SIMILARITY).  
 CC FT CONFLICT 64 64 D -> N (IN REF. 1).  
 CC FT CONFLICT 221 221 R -> G (IN REF. 1).  
 CC FT CONFLICT 242 242 R -> S (IN REF. 1).  
 CC FT CONFLICT 262 262 T -> G (IN REF. 1).  
 CC FT CONFLICT 345 345 S -> N (IN REF. 2).  
 CC FT CONFLICT 551 552 EL -> DV (IN REF. 3).  
 CC FT CONFLICT 688 688 D -> AI (IN REF. 3).  
 CC SEQUENCE 1026 AA; 117955 MW; 60852B647FE9057 CAC64;  
 Query Match 48.6%; Score 2626; DB 1; Length 1026;  
 Best Local Similarity 51.5%; Pred. No. 4.3e-159;  
 Matches 532; Conservative 177; Mismatches 301; Indels 24; Gaps 8;  
 QY 17 MESALTRDVRGVODFVLENTSEAFTELRARRRENILTYIGVLYSVNPRDLQI 76  
 DB 1 METGLHERDRAGVDFVILENTYSEAFTELRARRRENILTYIGVLYSVNPRDLQI 60  
 QY 77 YSRQHMERYGVSYEVPHLFAVADTVYRALRTERDQAVMISGSGAGKTATKRLQ 136  
 DB 61 YTDHVAAYRKNRHHYEMPHLFAVTDNAFRSLLEENGCQCVLLSGSGSKTASAKVLD 120  
 QY 137 FYAFETCAPERGVRRLLQSNVLEAFGNATLRRDSSRFGKTYWDQDFKGAAPVG 196  
 DB 1 FYAFETCAPERGVRRLLQSNVLEAFGNATLRRDSSRFGKTYWDQDFKGAAPVG 196



```

OY 207 RVVHONHGERNFHVYQLLEGCEETLRRLGLERNPOSTLYLVKGGCAVSSINDSK 266
DB 187 RVVKELEGERNFIYFOLLAGADAQLKALKLERDTGCAIYL-NPTSTVYDGDADDANK 245
OY 267 VMKALSVIDTFDEVEDDLSTVAVLHGNHFADEDSN----AQVTENQKLYTFL 322
DB 246 VLOSANTVIGFSEETROYEVAALVAKLGNVELINEFQANCVPASGIDGCGVEIGEL 305
OY 323 LGVECTTLREALTHRKLIKGEBLSPLNLEQAAVARDLAKAVSRPTLVKRLINSL 382
DB 306 VGLNSVELRALCSRTMETFAKEKVTTLNVAQYAROLDANLNTSRLENLVLNINESI 365
OY 383 ASKDDESPWRSTVYGLDITGFYFQHNSEPCINCYNEKLOQLFTELTKEQSEY 442
DB 366 KVGTEB-----KRYMGVLDIDIGFIELEDSNFEQVINYCNKLOQVLEMTLKEQEY 420
OY 443 ENEGIAEMVQYFNKKIICDLVEEFKGIISTLDEECRPGANTDTLEKLEDTVKRHP 502
DB 421 KREGIPWVKEVEYDNGIICNLLEHNRGLLAMLDEECRPGVSDSTFLAKLNOLFESKHS 480
OY 503 HPLTHKLADOKTRK---SLDGEFRLHYAGEVTVYVTFGLDKNDLFLRLKETMSSM 559
DB 401 HY-ESKVYQNMOKQYDHSMGLSCFRICHYAGKVYVNVNSFIDKNDLFLRLSQAMKAR 539
OY 560 NPIMACFDKSE--LSDKKREYATOPKMSULQVLELBSKEPARYRCIKPNDAKOPGR 617
DB 540 HPLRLSFLPEGDPKOASLKRPTTGAOKRSSVTTLMKMLXSKNPNYTRCIIKPNHQGRH 599
OY 618 FDEVLIHQVYKYLGLMENLRYRAGFAVRRKYEAFLOKYSKCPETPMAGRODDVAV 677
DB 600 FSEFLSVQAOYLGLLENVVRAGVAVRQVGSFLERYLLSTVPRMNGOQEVK 659
OY 678 LVRLHGYREPEYKMGRTKIFTRFKTLFATEDSLVBRQSLATKIOAMGFHROKFLR 737
DB 660 VLGELSSSEELAGKTKIFTRSEKTLFEEQRRRLQOALATLICKYRGWRCTRHYOL 719
OY 738 VKRSALCIQSWMRCTLGRKAKRMAAQTIRLRIGFILLRHSR-----CPENAF 788
DB 720 MRKSQIVISSWFRGNMOKKHRYRKMAKASALLQAVRQWKRKRKYFRSCAALLISNFI 779
OY 789 FLDHVRASFLINLRQLPR-NVLDTSWPTPP-----PALREASELLRELCKNMVKKYCR 842
DB 780 YKSNVO-KFLGLKLNDSPLSPSILDKKWPASARYFYFNANHELRLFIQ-----WK-CK 830
OY 843 ----SISPEKQOOLQKAVASEIFKCKNDYPOSVPLFSTLGTGHEISPRVLOLGSSE 898
DB 831 KFRQDQSPKQVEVLRKELCSELEFKGKASYPOSVPIPHGDTIGLQR-NPKLOKLGGE 889
OY 899 --PIQYAVPVYKTRDKGYKRRPOLLTPSAVVI--VEDAKVQRIDVANTLGISVSSLS 954
DB 890 EGPLLAETVYKVKRNKNAKTSRILLTGTGVIITDMKNPOAKTVIPLNSLAGVSTSRK 949
OY 955 DSLFLVHVOREDN-KQKGDVYVLOSDHVIETLRK---TALSADRVN-----NINIQ 1001
DB 950 DGLFSLHLSLSSVSGKEFLVSEHVIETLRKICRATLDAQOMLPVTVEEFSVKEKE 1009
OY 1002 GSIT--FAGPGRGD 1014
DB 1010 GSLTKVVIQPGGGG 1024

```

RESULT 5

MYIA\_MOUSE STANDARD: PRF: 1107 AA.

```

AC P46735: P70244:
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Myosin IA (Myosin I alpha) (MMI-alpha) (Mita) (Mih-L).
GN MYOIA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=C57BL/6; TISSUE=Cerebellum;
RX MEDLINE=93194946; PubMed=8449986;
RA Sherr E.H., Joyce M.P., Greene L.A.;
RT "Mammalian myosin I alpha, I beta, and I gamma: new widely expressed
genes of the myosin I family."
RL J. Cell Biol. 120:1405-1416(1993).
RN [2]
RP SEQUENCE OF 13-1107 FROM N.A. (ISOFORM 2).
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=93209370; PubMed=8458427;
RA Kozlovsky J.S., Qian C., Jiang X., Mercer J.A.;
RT "Molecular cloning of a mouse myosin I expressed in brain."
RL FBS Lett. 320:121-124(1993).
CC -1- FUNCTION: MOTOR PROTEIN THAT MAY PARTICIPATE IN PROCESS CRITICAL
TO NEURONAL DEVELOPMENT AND FUNCTION SUCH AS CELL MIGRATION,
NEURITE OUTGROWTH AND VESICULAR TRANSPORT.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: PROMINENT EXPRESSION IS SEEN IN THE BRAIN,
LUNG AND LIVER. IT IS ALSO EXPRESSED IN THE HEART AND TESTIS. A
HIGH LEVEL EXPRESSION IS SEEN IN VIRTUALLY ALL NEURONS (BUT NOT
GLIA) IN THE POSTNATAL AND ADULT MOUSE BRAIN AND IN NEUROBLASTS OF
THE CEREBELLAR EXTERNAL GRANULAR LAYER.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 5 IQ DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L00923; AAA39800.1; -.
DR EMBL: X69987; CAA49604.1; -.
DR HSSP: P08799; IAMD.
DR MGD: MGI:107732; Myo1a.
DR InterPro: IPR000048; IQ_region.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head.
DR Pfam: PF00612; IQ_5.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head.1.
DR SMART: SM00015; IQ_5.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS50096; IQ_4.
KW Myosin; ATP-binding; Actin-binding; Calmodulin-binding; Repeat;
KW Multigene family; Alternative splicing.
FT DOMAIN 704 727
FT DOMAIN 728 749
FT DOMAIN 750 778
FT DOMAIN 780 807
FT DOMAIN 808 837
FT NP_BIND 108 115
FT DOMAIN 592 599
FT VARSPIC 791 819
FT CONFLICT 44 44
FT CONFLICT 260 260
FT CONFLICT 304 304
FT CONFLICT 307 317
FT CONFLICT 333 333
FT CONFLICT 546 546
FT CONFLICT 570 571
FT CONFLICT 742 742
SQ
1107 AA: 128527 MW: FA244E5D2DE752A5 CRC64;

```

Query Match 38.7%; Score 2090.5; DB 1; Length 1107;  
 Best Local Similarity 41.7%; Pred. No. 5,1e-125;  
 Matches 463; Conservative 183; Mismatches 346; Indels 119; Gaps 19;

Oy	1008	G6PGRDGIIDFTSGSELLITKAKNHLVVA	1038
Dz	1083	QGNQNGSSVP-----TCRKNRLLEDA	1105
RESULT 6			
ID	MYIA_RAT	STANDARD:	PRT: 1136 AA.
AC	Q05096;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Myosin IA (Myosin I alpha) (MMI-alpha) (Mmia) (Myosin heavy chain myr 1)		
CN	MIOIA.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_Taxid=10116;		
RA	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Sprague-Dawley;		
RX	MEDLINE=93194945; PubMed=8449985;		
RA	Ruppert C., Kroschewski R., Baehler M.;		
RT	"Identification, characterization and cloning of myr 1, a mammalian myosin-I".		
RL	J. Cell Biol. 120:1393-1403(1993).		
CC	- FUNCTION: MOTOR PROTEIN THAT MAY PARTICIPATE IN PROCESS CRITICAL TO NEURONAL DEVELOPMENT AND FUNCTION SUCH AS CELL MIGRATION, NEURITE OUTGROWTH AND VESICULAR TRANSPORT (BY SIMILARITY).		
CC	- ALTERNATIVE PRODUCTS: 3 ISOFORMS: A (SHOWN HERE), B AND C; ARE PRODUCED BY ALTERNATIVE SPLICING.		
CC	- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.		
CC	- SIMILARITY: CONTAINS 5 IQ DOMAINS.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - between the European Bioinformatics institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/a or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL: X68199; CAA48287.1; .		
DR	HSSP: P08799; lMD.		
DR	InterPro: IPR000046; IQ_region.		
DR	InterPro: IPR001609; myosin_head.		
DR	Pfam: PF00063; myosin_head; 1.		
DR	Pfam: PF00612; IQ; 6.		
DR	PRINTS: PR00193; MYOSINHEAVY.		
DR	ProDom: PP000355; myosin_head; 1.		
DR	SMART: SM00015; IQ; 6.		
DR	SMART: SM00242; MYSC; 1.		
DR	PROSITE: PS50096; IQ; 5.		
KW	Myosin; ATP-binding; Actin-binding; Calmodulin-binding; Repeat; Multigene family; Alternative splicing.		
ET	DOMAIN 1 703 MYOSIN HEAD-LIKE.		
ET	DOMAIN 704 729 IQ 1.		
ET	DOMAIN 730 750 IQ 2.		
ET	DOMAIN 750 778 IQ 3.		
ET	DOMAIN 780 807 IQ 4.		
ET	DOMAIN 808 837 IQ 5.		
ET	DOMAIN 837 866 IQ 6.		
ET	NP_BIND 108 115 ACTIN-BINDING (POTENTIAL).		
ET	NP_BIND 592 599 ACTIN-BINDING (POTENTIAL).		
ET	VASAPLIC 794 823 MISSING (IN ISOFORM 1B).		
ET	VASAPLIC 794 852 MISSING (IN ISOFORM IC).		
SQ	SEQUENCE 1136 AA; 131917 MW; 69DC2A84E9070E2 CRC64;		
Query Match 38.5%; Score 2080; DB 1; Length 1136;			
Best Local Similarity 40.6%; Pred. No. 2.4e-124;			
Matches 463; Conservative 183; Mismatches 346; Indels 148; Gaps 19;			



```

QY 27 VGVODVLENFETSEAAFIENLRFRFRNLITYTIGVLVSNVPRDLOYSRQHMRYR 86
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 15 IGVODTVLEBPL-NEEFIDILKKRPHNEIYITIGSVIVSNPRSLPTISPEKVEDYR 73
QY 87 GVSFEVPRPHLFADVDVYVYRALKTERBDQAVMISGESGAGKTEATKRLLOFAETCPAPE 146
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 74 NRNYELSPHIFALSDAYRSRLRDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 133
QY 147 RGAVDRLQSNVLEAFNNAKTLNRDNRNRRFGKVDVDFDKAPVGGHILSTLLEKS 206
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 134 EVANQVRFQQLQSVPLVLEAFNNAKTLNRDNRNRRFGKVDVDFDKAPVGGHILSTLLEKS 193
QY 207 RVYHONIGENRNVYFQVLEBGEETLRRLGLERNPOSYLYLVKGCACAVSSINDKSDK 266
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 194 RVYKQPRGERNFVYFQVLEBGEETLRRLGLERNPOSYLYLVKGCACAVSSINDKSDK 252
QY 267 VMKRALSVIDFTEDEVEDLSTIYAVLHGNHFADEDSN-----AQVTENOLKYLTRL 322
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 253 TVRANQVGFSDPEAESVLEVAVALKGNIEFKPESRNGIDESKIDKNELKEICEL 312
QY 323 LGVEGTTLRALTRKRIIAKGEELSLPLNEQAAYARDAKAVYSRPTWLVKRNISRL 382
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 313 TSDIQVLERAFSFTVEAKOEKVSFTTLNVAQAYARDAKAVYSRPTWLVKRNISRL 372
QY 383 ASKDAESPMSRSTYVGLDIYGEFEVFOHNSFEQCIANCNEKLOQLFIETLKSEOEY 442
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 373 KQOTVRRK-----VAGVLDITGFELFEDNSFEQCIANCNEKLOQLFIETLKSEOEY 427
QY 443 EMBGLAMEVVOYFNKKIICDLVEERFKGIISILDEECRLPGCATDFTLEKLEDVYKHP 502
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 428 IREDIEMHIDYFNNAIICDLLENMTNGLAMLEDEECRLPGCATDFTLEKLEDVYKHP 487
QY 503 HFLTH--KLADQTKRKSIDRGEFLNHYAGVYTSVTFPLKNDLLEFRNLKETWSSMN 560
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 488 HESMSKCSRLNLTTLPHSCFRIQHYAGVLYQVEGFVKNNDLTYRDLQAMMAGH 547
QY 561 PTMAOCFPKSELS--DKRPTVATQFMSLSLQVLEIRSKPEYVIRIKRNDKAKQGRF 618
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 548 ALIKSLFEGCNPAKVNLRPTASQGFASVATLMLNQTNPVYIRIKRNDKAKAHIF 607
QY 619 DEVLIRHOVKLGLMENLRVRAGFAVRRKYEAFRLQRYKSLCPETWPMAGRPDQVAVL 678
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 608 SESLQCHQIRYGLLENVRAGYAFQAYEPCLERYKMLCKQTPWHMKGAPASGEVL 667
QY 679 VRHLGYKDEEYKMGRTKFIFFPKTLFATBEDSLEVRQSLATKIQAMRGEFHMQRKLV 738
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 668 FNELEIPEEYSEFGRSKIFIRNPRTLQLEDLKKORLEDAFLQKLYKGMKQRTHELM 727
QY 739 KRSATCQSMWRGTLGRRAKAKRMAAQTIRLRGF-----ILRSHRCPENA---- 787
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 728 KRSQVIAAMVRRYAOQRYOQIKSSALVIOYIRGMKARKILRELKHOKRCEAATTIA 787
QY 788 ----- 787
Db 788 AYWHGTQARKERRRLKDEARNKNAIYVAFWLSGARRELKRLKEARRRHNAVAVIWAY 847
QY 788 ----- 787
Db 848 WLGLKVARREYRFFRANAKKIYEFLLQRYQVLYLEMMKNKMSIPIDKNMPSRYLFL 907
QY 823 EASELLRELCKMNMV-----KYCRSISPEKKQOLQOKAVASELTFKGGKNDYPOSVPLRFTS 879
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 908 DSHH--KELKRIFHLWRCKKRYRQDQFDOOKLYIEEKLASELSEFKDKALYPSVQCPFG 965
QY 880 TRLGTEISB---RVLQSGEPTIOYAVPVYKDKGKPPRROLTLPPSVVIVE--D 933
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 966 AVL---EINKNPKYKYLKDAIEKIIIAEVYKINANKSKSRIFLLNNLLADOKS 1022
QY 934 AKYKQIDVANIJGISVSSLSFLVHQR--EDNKQKQDVVLQSDHVTETLK--TAL 989
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1023 GQIKSEVPLVADYTKVSMSSQNDGFFAVHLKEGSEAAKGFLESSDHLLEMTATLYRTLL 1082
QY 990 SADRYN-NINI-----NQSITFAGCGRCGCIIDFTSGSELLTTKAKNGHLAAYA 1038

```

```

Db 1083 SOTKOKINIEISDFLEQVPRQDKVCYKRFIQGNQKNGSVP-----TCRKNRNLLEVA 1134
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 7
MYHL_HUMAN
ID MYHL_HUMAN STANDARD; PRT; 1043 AA.
AC Q9UDC5; Q9UD7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Brush border myosin I (BBM-I) (Myosin I heavy chain) (MIHC).
GN MYHL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RA Biale D.D., Munson S.J.;
RT "Human brush border myosin I.";
RN Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Jejunal brush border;
RX MEDLINE=2011970; PubMed=10653589;
RA Li W., Wang J., Coluccio L.M., Matsudaira P., Grand R.J.;
RT "Brush border myosin I (BBMI): a basally localized transcript in human
RL jejunal enterocytes.";
RN J. Histochem. Cytochem. 48:89-94(2000).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99073867; PubMed=9858156;
RA Skowron J.F., Bement W.M., Mosseker M.S.;
RT "Human brush border myosin-I and myosin-Ic expression in human
RL intestine and Caco-2BBE cells.";
RN Cell Motil. Cytoskeleton 41:308-324(1998).
-1- FUNCTION: INVOLVED IN DIRECTING THE MOVEMENT OF ORGANELLES ALONG
CC ACTIN FILAMENTS (POTENTIAL).
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 IQ DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF127026; AAC31189.1; -
DR EMBL: AF105424; AAC78645.1; -
DR EMBL: AF009961; AAC27457.1; -
DR HSSP: P08799; IMND.
DR Genew: HGNC:7595; MYO1A.
DR InterPro: IPR000048; IQ_region.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head; 1.
DR Pfam: PF00612; IQ; 3.
DR PRINTS: PD00193; MYOSINHEAVY.
DR PRODOM: PD000335; myosin_head; 1.
DR SMART: SM00015; IQ; 3.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS50096; IQ; 2.
KW Myosin; Actin-binding; ATP-binding; Calmodulin-binding; Repeat;
KW Multigene family.
FT DOMAIN 1 681 MYOSIN HEAD-LIKE.
FT DOMAIN 697 719 IQ 1.
FT DOMAIN 720 742 IQ 2.
FT DOMAIN 743 772 IQ 3.
FT NP_BIND 101 108 ATP (POTENTIAL).
FT DOMAIN 571 593 ACTIN-BINDING (POTENTIAL).

```



FT DOMAIN 746 774 10.3.  
 FT NP\_BIND 104 111 ATP (POTENTIAL).  
 FT DOMAIN 574 596 ACTIN-BINDING (POTENTIAL).  
 FT CONFLICT 46 46 E -> O (IN REF. 2).  
 FT CONFLICT 285 285 C -> S (IN REF. 2).  
 FT CONFLICT 297 297 T -> A (IN REF. 2).  
 FT CONFLICT 427 427 G -> A (IN REF. 2).  
 FT CONFLICT 852 855 CASE -> WPRO (IN REF. 2).  
 SQ SEQUENCE 1045 AA; 119276 MW; 4C688BCA3FD669D0 CRC64;

Query Match 37.2%; Score 2005.5; DB 1; Length 1045;  
 Best Local Similarity 41.9%; Pred. No. 1.2e-119;  
 Matches 438; Conservative 197; Mismatches 375; Indels 35; Gaps 17;

QY 17 MESALTARDVGVODEFVLENTSEAFIENLRFRFENLITTYTIGPLVSVNPRDIOI 76  
 DB 1 MEATSLDAAVAGDVLMDLP-SEESLRLTQERFSGELYITIGEVISVNPYKPLPI 59  
 QY YSRQIMERYRGVSFEYEPVPHLEPAVDIVYKALRTERDQAVMISGSCAGTEATKRLLO 136  
 DB 60 YTPKEVEYHNCNFEVAVPHTYAIADDAVRSRLRDRDRDQICILLIGESAGCKEASKLWMS 119  
 QY 137 FYAETCPARGCAVRODLQSNVYLEAFNAKTLRNDNSRFGKYMDVDFEGAVG 196  
 DB 120 YVAVSSKGEVDKVEEDLQSNVYLEAFNAKTLRNDNSRFGKYMDVDFEGAVG 179  
 QY 197 HILSYLLEKSNVHQNHEERNEHVFYOLLEGEETLRRLGLEPNQSYLYLVKGCACV 256  
 DB 180 VISNLLKSRIVRHVGERNFHIFYOLLGSGAQLQOLKLRDCHGYTL-NHEKSVL 238  
 QY 257 SSINDSPMKYRKALSYIDTDEVEDLSIVASVHLGNH---FAADEDNSNAOVT 312  
 DB 239 PGMDAAFRMADAMALIGFAPEAVTALLEVAVLKLONVKSQCFQASGMEASSITE 298  
 QY 313 ENQKLYLRLGLVGTTLREALTRKTIAGCELLSPLEQAAVADALAKAVSRFTF 372  
 DB 299 PREQIEISQILGLDPSLTLEQALCRITKAVNDESLTSLVSGYTGADALAKNISRFLD 358  
 QY 373 WLVAKIRSLASKDAESPNSRTTVLGLDIYGEVFOHNSPEQFCINYNENKILQDLFE 432  
 DB 359 WLVRNITTSIQVPG---KORVMGVLDIYGEFIFQDNGFEQFIINYNENKILQDLFE 413  
 QY 433 LTLKSEDEEYEAAGIAMEPYOVFNKKIICDLEVEKFGITSLIDECLRGENTDLFLE 492  
 DB 414 MTLKEDEEVYRGIOKTPYEFEDNSIICLIENSKVGIAMDECLRGVNEDEFTT 473  
 QY 493 KLEDYKPHPHETLTKLADQ--TRKSLDGEFRLHYAGEVTVSYTGFELDKNNDLFFRN 550  
 DB 474 KLMQIFASHRYESKFTLNKAHVTDSLPLRCFRHHYAGKVYVNTGFIENKNDLFFRD 533  
 QY 551 LKFTMGSSNMPIMAGCFDS--ELSKKREYVATOPKSLDQVLETLRSKEPARYRCIK 608  
 DB 534 LSGAMMAARITTLRLSPLEPDDPDRPSLKPPTTGSQFKASVATLMKLNLSKNPNYTRCIR 593  
 QY 609 PNDAPGPREDEVLLRHQVAYLGLMENLHVRRAGFAVRRKYEAFLORYSLCPETPMMA 668  
 DB 594 PNDTKTAMLETPDLVLAQVRYGLMENAVRRAGVAFRLQYLPFLERYMLSKRTWPKMT 653  
 QY 669 GRPQDQVAVLVRLGKPREPKMGKRTKIPRPTKLPATEDSLEVRQSLATKIOAAMG 728  
 DB 654 GGRGAEVALLLEKPPPELAYGHTKIFIRSPRTLFDEKRRQORVALATLIQKMFGR 713  
 QY 729 FIMRQKFLVAKSAICIOSMMRGTLGRKAARKMAAOTIRLIGELIR-----HSP 781  
 DB 714 WCCRKRYQIMRKSQLISIMRFGHMORNRKYOKMRSVLLQVYAGKWTFRMYRYFRSD 773  
 QY 782 RQPE-NAFLDIVASFELNLRRL-PRNVLDTSWPTTP-PALREASELRELCKKNVW 838  
 DB 774 ACTRLSNFYRRMVQKYLIMGLKNLPMAVLDRTWPPAPYKFLSDANGLSIFRKMCK 833  
 QY 839 KYCRSISPEMKOOLQKAVASEIFGKKDNYQSVPRPLFISTRGL-TEEISPRVLOSIGS 897  
 DB 834 KYRDLTPOQRAMLOAKLCASELFDKRAKALYAQSLQDPFREGYLGTLTONRRYOKLOAVAK 893

QY 898 EPIQYAVPVYKDYKGRPPRROLLPNSAVVED--AKKORIDYANLTGISVSLSD 955  
 DB 894 KILVMAEAVQKVNRAKTVPRILLITTEHLVADRAKQPKNVLSLQIQGASVRSRD 953  
 QY 956 SLFVLYHQREDNK-QKRDVYLOSDDHYETITK---TALSAD-RVNNINI-NGOSTEPAG 1009  
 DB 954 GLALHUKETSTYAGKDDLLVSPHLELVTRLHQRLMATQALPISADQFSTRFPKG 1013  
 QY 1010 PGRDGIIDFT-SCSELLITKAKNGH 1033  
 DB 1014 DVAVTVESAKGGGDVPCVCKKRSH 1038

RESULT 9  
 MYHL\_MOUSE  
 ID MYHL\_MOUSE STANDARD; PRT; 909 AA.  
 AC C88329;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Brush border myosin I (BBM-I) (BBMI) (Myosin I heavy chain) (MHC)  
 DE (Fragment).  
 GN MYHL OR BBM-I.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99132700; PubMed=9933937;  
 RA Skowron J.F., Mosseker M.S.;  
 RT "Cloning and characterization of mouse brush border myosin-I in adult  
 and embryonic intestine."  
 RL J. Exp. Zool. 283:242-257(1999).  
 CC -I- FUNCTION: INVOLVED IN DIRECTING THE MOVEMENT OF ORGANELLES ALONG  
 ACTIN FILAMENTS (POTENTIAL).  
 CC -I- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF009960; AAC28397.1; -  
 DR HSSP; P08799; 1MND.  
 DR MCD; MG1:1338019; Myhl.  
 DR InterPro; IP000048; IO\_region.  
 DR InterPro; IP0001609; myosin\_head.  
 DR Pfam; PF00063; myosin\_head; 1.  
 DR Pfam; PF00612; IO; 3.  
 DR PRINTS; PR00193; MYOSINHEAVY.  
 DR ProDom; PD000355; myosin\_head; 1.  
 DR SMART; SM00015; IO; 3.  
 DR SMART; SM00242; MYSC; 1.  
 DR PROSITE; PS50096; IO; 2.  
 KW Myosin; Actin-binding; ATP-binding; Calmodulin-binding; Repeat;  
 KW Multigene family.  
 FT DOMAIN 1 681 MYOSIN HEAD-LIKE.  
 FT DOMAIN 697 719 IO 1.  
 FT DOMAIN 720 742 IO 2.  
 FT DOMAIN 743 772 IO 3.  
 FT NP\_BIND 101 108 ATP (BY SIMILARITY).  
 FT DOMAIN 571 593 ACTIN-BINDING (POTENTIAL).  
 FT NON\_TER 909  
 SQ SEQUENCE 909 AA; 104421 MW; F7EF392D5837341 CRC64;  
 Query Match 36.8%; Score 1987; DB 1; Length 909;  
 Best Local Similarity 44.8%; Pred. No. 1.4e-118;

Matches 406; Conservative 177; Mismatches 292; Indels 32; Gaps 12;

QY	27	UGVODFVLNPFNSSEAFINLRRREPRENLIYAYIGVILVSNVYRDLQIYSQNHMYR	86
Dd	8	VGVEDLLILEPL--DEESLILNLOLRTEKENKLTYYIGVILVSNVYRDLQIYGEFELAKTR	66
QY	87	GVSEFEVPEPHLFVADVYVYRALRTERRDQAVMISGESGACKTEATKRLLOFVAETCPAPE	146
Dd	67	DYFFELAKPHIYALANAYASLOMDRODQOLILITGESGACKTRXXKILVMSYVAVCGKE	126
QY	147	RGCAVDRLLQSNVLEAFGNAKTLRDNSSRGKTYWDOVDFKGAPVGHILSYLLEKS	206
Dd	127	QVNSVKEOLLQSNVLEAFGNAKTLRNNSSRGKYMDIEFDEKSPFLGVTINYLLEKS	186
QY	207	RVYHONHGEENHNFYOLLLEGGSEETRLGLDENPQSYLYLVKGCQAKSSINDSDMK	266
Dd	187	RYVAKQLGRPFHLEFYOLLGADQOLKALKLEEDTISVYTL--NGEYSKNGMDASNFR	245
QY	267	VMRKALSVIDTBEVEDLLSTIVASVYLHGINHFADEDSN---AQVTTENOILKYTLRL	322
Dd	246	AVQNAHMSVIGRSEELINQVLEVALVALKGNVKTLDPEQANGIPASIDCGGKQIOLGEM	305
QY	323	LGVEGTTLRREALTHRKITIAAGEELISPLNLEQAAVARDALAKAVYSTFPWLVKIRNRL	382
Dd	306	MGJNSTELERLCSRTMETGEKEKVYVLTANOAYARDALAKINYSRFLPMIVKIRINESI	365
QY	383	ASRDAESPMSRSTVLLDILITIGEVYOHNSRFOFCINYNCEKIOOLFIELTLKSDEEY	442
Dd	366	KVGTGEKK----VMGVLDIYGEELIEDNSFBOVLYNTYNERLOOVIELTLKEDEEY	420
QY	443	EAEIGIAPVOYFNNKICIDLVKEKFKGISILDEECLRGGEATDLFLEKLEDETVRPH	502
Dd	421	KREGIPMTKVEYFNGJICNLIEHSQGLIAMDDEBLRGCVVSDSRFLKLNQLEPSKS	480
QY	503	HELTHLKLAQKTRKSLDR---GEFLHTAGEVYTSVTFGLDKNNDLFRNLKETWCSS	558
Dd	481	HY--ESKVSQNAQOXYDRTMGLSCFRISHYAGKVTYVNTGFIIDKNNDLFRDLSQTMWKA	538
QY	559	MNPIMACCFKJSELSDK--KREPVIATOPKMSLQIOLVIELRSKPEATIRCTKPRDAQOP	616
Dd	539	QHPLKLSLFPREGNPKESLKRPPAGIOFNSVAVLNKLNKYSKPNYIRCTIKPRDDQOKG	598
QY	617	RFDEVLIRHOVKYUGLMBNLVRRRAGFAFRKRYEAFIORYKSLCEPTWPMWAGRPQOVA	676
Dd	599	RFTSEMMVQARYTGJLLENVVRAGAFANQGYKFRPELERTRLRSRSTWPMKNGDDRGVE	658
QY	677	VLVNHLGKPEEYVMGRTKIFIRPKTLFATEDSLEVYROSLATKIOAMRGFHMROKFL	736
Dd	659	KVLSTLSTSEELIYAGTKTIFIRSPKTLFYLEBQRLRLQOALATLQKVYRGMCRPTHYQ	718
QY	737	RKRSALCISQSMWGTGLGRKKAARKKMAAOTIRLILNGFILHRSRPRENA-----F	788
Dd	719	QMRKSQILISMPFGNKKOKKHGKIRSSVLLIOQFVGMWRKRRKYRYFRSGAALLTANF	778
QY	789	FLDHFVRSFLNLRRLP--RVNLDTSWPTP--PALRASELRLRELCKNNMYMKCRSISP	846
Dd	779	IYQSMAGKFLNLKKNLPSTVYLDNTMPAPYRCOFNPNANGLOLFLYQMKCKFRDOLSP	838
QY	847	EMKQOLOOKAVASEIFKQKDNVPOSVPRLEFISTRIGTEETISPRVLOSLSGE--PIQAV	904
Dd	839	KOYOTLREKLCASELFGKKASYPQSVIPFRDYGIGQ--NPKQLGKLGREBPVLVAD	897
QY	905	PVAYKDR 911	
Dd	898	TVKKVNR 904	

```

DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bruch border myosin I (BBM-I) (BBMI) (Myosin I heavy chain) (M1HC)
DE (Fragment).
GN MYHL OR BBMI.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver, and Testis;
RX MEDLINE=9528044; Pubmed=7779104;
RA Balish M.F., Colucio L.M.;
RL "Identification of brush border myosin-I in liver and testis.";
CC CC
CC -1- FUNCTION: INVOLVED IN DIRECTING THE MOVEMENT OF ORGANELLES ALONG
CC ACTIN FILAMENTS (POTENTIAL).
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 IO DOMAINS.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U25148; AAA89132.1; -.
DR HSSP; P08799; IMND.
DR InterPro; IPRO00048; IO_region.
DR InterPro; IPRO01609; myosin_head.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00612; IO; 3.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRODom; PD000355; myosin_head; 1.
DR SMART; SM00015; IO; 3.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS50096; IO; 2.
KW Myosin; Actin-binding; ATP-binding; Calmodulin-binding; Repeat;
KW Multigene family.
KW NON_TER 1
FT DOMAIN <1 673 MYOSIN HEAD-LIKE.
FT DOMAIN 689 712 IO 1.
FT DOMAIN 713 733 IO 2.
FT DOMAIN 735 764 IO 3.
FT NON_TER 842 842
SQ SEQUENCE 842 AA; 97210 MW; DDAFDCCB6308316B CRC64;

```

Query Match	35.8%;	Score 1932.5;	DB 1;	Length 842;
Best Local Similarity	45.6%;	Pred. No. 3.7e-115;		
Matches 388;	Conservative 165;	Mismatches 269;	Indels 29;	Gaps 10;

[illegible]



```

DR Pfam: PF00063; myosin_head. 1.
DR Pfam: PF00612; IQ. 2.
DR PRINTS: PR00193; MYOSINHEAVY.
DR PRODOM: PD000355; myosin_head. 1.
DR SMART: SM00015; IQ. 1.
DR SMART: SM00242; MYSC. 1.
DR PROSITE: PS50096; IQ. 1.
KW Myosin: ATP-binding; Actin-binding; Calmodulin-binding; Repeat;
KW Multigene family.
FT DOMAIN 1 694 MYOSIN HEAD-LIKE.
FT DOMAIN 695 721 IQ 1.
FT DOMAIN 716 736 IQ 2.
FT NP_BIND 100 107 ATP (BY SIMILARITY).
FT DOMAIN 567 589 ACTIN-BINDING (BY SIMILARITY).
SQ SEQUENCE 1011 AA; 117094 MW; 31206C065B5D5DFA CRC64;

Query March 33.48; Score 1800.5; DB 1; Length 1011;
Best Local Similarity 39.68; Pred. No. 1.2e-106;
Matches 426; Conservative 169; Mismatches 341; Indels 139; Gaps 29;

QY 28 GVQDFVLLNFTSEAFETENLRFRRENLIYTYIGPVLVSNPYRDLQIYSRQHMERYG 87
DB 8 GVQDFVLLDQYSME-KFMDNLKRFRQNSITYTYIGEVCSNPNYQNMITYGETIRKKG 66
QY 88 VSFTEVPPLFAVDYTRALRTERRQDAVMISGESGAKTEATKRLQFAETCPAP-- 145
DB 67 RELFENAPHLPAIDSAVRVKORODTCILISGESGAKTEASKIKMKYIAAVTNAGQ 126
QY 146 ---ERGGAVRRLLQSNVLEAFNAKTLRNDSSRFKQYDVQDFGAPVGGHLSYL 202
DB 127 NEIER---VKNVLIQSNALILEFGNAKTRNDSSRFKQYDIEDYKADPVGGITTYVL 183
QY 203 LEKSHVYHQNNGERNFHFVYQLEEGEETLRGLERNPQSYLVKQCAKVSINDK 262
DB 184 LEKSHVYQGGGERHGFHFYQLEGRANDNELRQYELQKETGKHHLNG---SMQILREK 240
QY 263 SDMKVKKRALSVIDTDEVEDLSTIVASVLHGNHRAEDSDSAQYTTNOLKYLTRL 322
DB 241 SDYKTCNAFTLIGSTDEVQTIWRTIAAVLHGNVEQTED-ELVISNKHQHLSTAKL 299
QY 323 LGVECTRLREALTHHKILAKGEBELSLPLNDAQAAARALAKAVYSRFTMLVKRINS 382
DB 300 LGVETELSTLTKRVIAAGVNMOKDHNAQAEYKADKALKAIDRLFTWITISINAI 359
QY 383 ASKDAESMSRSTVTLGLDLYGEFEVFOHNSFEPCINVCNEKLOQLTELTLEKEOEY 442
DB 360 LFRGSKTQA-RFNSYIGVDLYGFEIIPDSNFSFQRCINVCNEKLOQLTELTLEKEOEY 418
QY 443 EAEGIAVEPVYFNKKIICDLVEEKFKGIISLDECLRPGFATDLEFLKLEDTVPHP 502
DB 419 QREGIEMTNIIEVFNNKICLDLVEQPHKGIIMDEACLSVGKVTDDTLGAMDKNLKHP 478
QY 503 HFLTHKLADQKTRSL-DRGEFRLHAGEVYTSYTGFLDKNNLLFR-----NLKET 554
DB 479 HYSKQTL--KPLDKELKHREDFRITHYAGDYLYNNGEIEKKNKDLXODFKRLHNSDA 536
QY 555 MCSANNPAAOCFDESELSDDKKRPETVATOPFMSLQLEIIRSEKQYIRCIKPNDAK 614
DB 537 NLSEKMPGAGADIKT-----TKRPLTACTLQFORSMAADVLLTKKEPYVRKICAKNDKS 592
QY 615 PGREDEVLIHQVLYGLMELRVRRAGFAYRKYAEFLQRYKSLCPETWPMW-AGRPOD 673
DB 593 STVDEEVEHQVRYGLLENLRYRRAGFVHRQRYDKFLTRYKMSQYTWPFRAQSDRD 652
QY 674 GVAIVLVRHLGKPEYKKGRTKIFIRPKTFLATEDSLVRRQSLATYIQAAWGFHHRQ 733
DB 653 GVAIVLIEKKR-ADQVKGKHKIFIRSPRTFALHQRNEMIPHTVTLLQKVRKGIWVR 711
QY 734 KFLRVKSAICIQSWNRGTLGRKRAKRRKMAQOTIRLIRGLFIRHSPRCENAFELDHV 793
DB 712 NFKKKK-----AAITVRAVKAYKLR----- 732
QY 794 RASFLNLRRL--PRNVLD---TSWPTPPALREASELLR--ELCKMKNV-WKYGCS 843

```

```

DB 733 --SYVQELANRLRRAKQMRDQKSIQWQPPPLAGKRYEAKLHRMDFWRANMILHKYRS 790
QY 844 ISPEWKOOLQOKAVASELFKKCKNDYPOSVPRLFTSTGTEISPRVLQSLGS----- 897
DB 791 ---EW-POLRIQIITATLACGRPPWGA--RRWGDVLANSQENSGEATYNGSIKRN 844
QY 898 -----EPIQYAVPVYKDRKGYKPPRRLTLPPSAVIVEDAK-----VKQRIDYAN 944
DB 845 HPADETFQVLFSSFYKKNFHN-KQANRAFIYVSDSTLHKLDGKKNKFKDKRTIKIRE 903
QY 945 LTGISVSSLSPLVLAHQREDNKGQDV--LQSDHYIEFLTKALSADRYN----- 995
DB 904 LTSISVSGRDLIVFH-----SSKNK-DLVFSLESEY-----TPLKEDRIGEVVGIVC 951
QY 996 -----NININGQSIFFGGRGDIIDFSSGSELLITKAKNGHLAVAP 1039
DB 952 KKHDLTETELRVNVTNISCHLDGKARITTYEASNVENFRKESGNIIFEV 1006

RESULT 12
MYSE_DICDI
ID MYSE_DICDI STANDARD: PRT: 1003 AA.
AC 003479;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Myosin IE heavy chain.
GN MYOE OR DMIE.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AX3;
RC MEDLINE=93277957; PubMed=8504170;
RA Urrutia R., Jung G., Hammer J.A. III;
RT "The Dictyostelium myosin IE heavy chain gene encodes a truncated isoform that lacks sequences corresponding to the actin binding site in the tail";
RT Biochim. Biophys. Acta 1173:225-229(1993).
RL
CC -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE ACTIVITY THAT IS ACTIVATED BY ACTIN. MAY PLAY A ROLE IN MOVING MEMBRANES RELATIVE TO ACTIN.
CC -!- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED. DIMER OF A HEAVY AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 IQ DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
CC -----
CC EMBL: I06805; AAA33201.1; -.
CC PIR: S33760; S33760.
CC HSSP: P08799; IAMD.
CC DictyDb: DD01049; MYOE.
CC InterPro: IPR000048; IQ_region.
CC InterPro: IPR001609; myosin_head.
CC Pfam: PF00063; myosin_head. 1.
CC Pfam: PF00612; IQ. 2.
CC PRINTS: PR00193; MYOSINHEAVY.
CC PRODOM: PD000355; myosin_head. 1.
CC SMART: SM00015; IQ. 1.
CC SMART: SM00242; MYSC. 1.
CC PROSITE: PS50096; IQ. 2.
KW Myosin: Actin-binding; ATP-binding; Calmodulin-binding; Repeat;
KW Multigene family.
FT DOMAIN 1 693 MYOSIN HEAD-LIKE.

```



```

KM Myosin: ATP-binding; Actin-binding; Calmodulin-binding; SH3 domain;
KM Multigene family.
FT DOMAIN 1 677 MYOSIN HEAD-LIKE.
FT DOMAIN 697 722 IO.
FT DOMAIN 1042 1099 SH3.
FT DOMAIN 579 589 ACTIN-BINDING (POTENTIAL).
FT NP_BIND 110 117 ATP (POTENTIAL).
SQ SEQUENCE 1099 AA; 125947 MW; 7D11592310C0C8F0 CRC64;

Query Match 32.0%; Score 1726.5; DB 1; Length 1099;
Best Local Similarity 38.0%; Pred. No. 6.7e-102;
Matches 399; Conservative 159; Mismatches 297; Indels 195; Gaps 24;

QY 28 GVQDFVLENTSEAFTELERRRENLIYTYIGPVLSVNPVRLQIYRQHMERYG 87
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 18 GVDMDVLLSQIT-EDAIYRNLRHKKRMDYIFTYISLVISVNPQMPYFTDRELDVQ 76
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 88 VSFYVPHLAVADTVYRALTERDDQAVMTSGSGAGKTEATRLLOFAETCPAPER 147
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 77 AVQYEPHIALYALTDNMYRNMLIDENCVLISGSGAGKTVAAKYIMQYISKVSGGDK 136
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 148 GGAVERDLLOSPLVEAFGNAKTLRNDSSRFGKMDVQDFDKAPVGGHLLSTLEKSR 207
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 137 VOHVADIILOSPLLEAFGNMAKTVRNNSRFGKTFEIOFSRGEGDGGKISNPLEKSR 196
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 208 VVQNHGERNFHYVQLLEGGEETLRRLGLERNPOSYLYLVKGOCAYSSINDKDMKV 267
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 197 VVQNHGERNFHYVQLLEGASQEQQNLGI-MSQDYIYLLQSDPYKEGDDSDPSE 255
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 268 MKKALSVIDFTEDEVDLSTYASVLHGLNHFPADEDSNAOVTENOLKYLTRLLGVEG 327
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 256 TLSAQMVGFRACQOOLVQLVAGLHGLNISF-CEEGYVAVESDSLAPFAYLLGIDS 314
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 328 TFLREALFHRKIIAK-----GELLSPNLLEQAAVARDALAKVSRPTFLVRKINRSLA 383
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 315 GRLQEKLSRKNDKSGRSESDVTLNVEQAKYTRDALAKGLYRDLDFLEALNRAV- 373
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 384 SKDASPSNRSTTVGLLDIYGFVEFOHNSPFCINCNEKLOQLTELTLKSEQEYE 443
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 374 QKPOEYS-----IGVLDIYGFELFOKNGFQFCINFNEKLOQLTELTLKAQEEYV 427
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 444 AGCIAMEPVQYNNKTIICDLYVEERK--GIISILDEBL--RPEBANDLFLKLEDTV 498
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 428 OGCIWMTPEYENNNKIVCDLIENKLSPPGIMSVLDVCAVTMATGGADQTLQKLOAAV 487
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 499 KHPHPLTKLADQKTRKSLDGEFRLLYAGEVTVSYVGFGLKNDLLEFRLKETMCS 558
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 488 GTHEHF-----NSWSAG-FVHHYAGKVSIVDSGFEERNRMDVLFSDLIELMOSS 535
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 559 NNPIMAOGF-DKSELSDDKRPETVATOFKMSLLOLVEILRSKRPAYIRCIKPNDAKQGR 617
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 536 DDDFLRMLFPERKLNIDKKGRPSTAGSKIKQANDVSTLAKCTPIHYIRIKRNEKRRPD 595
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 618 FEEVLIRHOVKTLGLMENLRVAGAPARRKYEAFLQRYKSLCPTWPMAGRPDQAVAV 677
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 596 WEESRYKHQVEVLEGRNIRVRAGFAFYRRQFSKFLQRAILLPTWPMWRGDEGQVH 655
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 678 LVRHLYGKPEEKMGRTKFIREFPTLFATETDSLEVRQSLTAKIOAARGHMQKFLR 737
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 656 LIRAVNMEPDQYOMOSTKVFANPSLFLLEMRKKEKDFGARTTOKAA----- 704
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 738 VKRSAICIOSMWRGTLGRKKAARKMAQOTIRRLIRGFIILRHSPPCPENAFPLDHYRAS 797
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 705 -----RRHVAVRKY----- 713
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 798 LILNRQLPRNVLDTSWTPPPALREASLLRELCKMKNWVKYCSISEWKQOLOQKAV 857
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 714 -----EEMRE----- 719
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 858 ASEIFKGGKNDYPOSVPLFISTRIGTEIISPRVLOSLS-EPLOYAVPVYKDYDRGKYP 916
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 720 ASNILLNKERRRRNSIRNFVGNVYGLBE-REPLKQFLAKRRERVDPAADSVTKYDRR-FKP 777
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

QY 917 RPRQLLTPSAVIVEDAKV-----KORIDYANLTGISVSLSDSLFVLHVQ 963
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 778 IKRDLITPKCYVYIGTEKRGKPEGLVREVLKRRKLDIQAALRGVSLSTRDDFFIL-Q 835
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 964 REDNNQKGDVVLQSDHIVETLTKT---ALSADRVNNINNGSISF-----A 1007
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 836 EE-----AADTFLESIFKTEFVSLCKRFEEDARRPLRPLFDILOFRVKKKGW 885
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 1008 GGPGRGIIIDFTSGS-ELLITKAKNGHLAV 1036
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 886 GGCTRN--VTFSGOVNLAVLKAGGKTTLTI 913
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 14
MYSB_DICDI
ID MYSB_DICDI STANDARD: PRT: 1111 AA.
AC P34092.1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin IB heavy chain.
GN MYOA OR DKIB.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=89345628; PubMed=2762320;
RA Jung G., Saxe C.L. III, Kimmel A.R., Hammer J.A. III;
RT "Dictyostelium discoideum contains a gene encoding a myosin I heavy
   chain";
RL Proc. Natl. Acad. Sci. U.S.A. 86:6186-6190(1989).
RN [2]
RP SEQUENCE OF 481-490; 656-666 AND 783-798.
RC STRAIN=AX3;
RX MEDLINE=93315475; PubMed=8325874;
RA Jung G., Fukui Y., Martin B., Hammer J.A. III;
RT "Sequence, expression pattern, intracellular localization, and
   targeted disruption of the Dictyostelium myosin ID heavy chain
   isoform.";
RL J. Biol. Chem. 268:14981-14990(1993).
CC -I- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
   ACTIVITY THAT IS ACTIVATED BY ACTIN. MYOSIN IB MAY HAVE A ROLE IN
   CHEMOTAXIS AND AGGREGATION. IT COULD SERVE TO STABILIZE AND EVEN
   RETRACT CORTICAL STRUCTURES, SUCH AS PSEUDOPODS AND LAMELLOPODS.
CC -I- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED. DIMER OF A HEAVY
   AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.
CC -I- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION JUST BENEATH THE
   PLASMA MEMBRANE IN THE ANTERIOR PSEUDOPOD AT THE LEADING EDGE OF
   THE CELL.
CC -I- DOMAIN: TH.1 BINDS DIRECTLY TO ANIONIC PHOSPHOLIPID MEMBRANES;
   MYOSINS I COULD THEREFORE MOVE ACTIN RELATIVE TO MEMBRANES AND
   VICE VERSA. TH.2 AND SH3 BIND TIGHTLY TO F-ACTIN. THIS TOGETHER
   WITH THE NUCLEOTIDE-SENSITIVE SITE IN THE HEAD, ALLOWS SINGLE
   MOLECULES OF MYOSIN I TO CROSS-LINK ACTIN FILAMENTS.
CC -I- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (see http://www.isb-sdb.ch/announce/
   or send an email to license@isb-sdb.ch).
CC EMBL: M26037; AAA33229.1; -.
CC PIR: A33284; A33284.
CC HSSP: P08799; 1AMD.
CC DictyDB; DD01047; myOA.
CC InterPro; IPR000048; IQ_region.
CC InterPro; IPR001452; SH3.

```



```

DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00018; SH3.1.
DR Pfam: PF00063; myosin_head.1.
DR Pfam: PF00612; IQ.1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR PRINTS: PR00452; SH3DOMAIN.
DR ProDom: PD000066; SH3.1.
DR ProDom: PD000355; myosin_head.1.
DR SMART: SM00242; MYSC.1.
DR SMART: SM00326; SH3.1.
DR PROSITE: PS50002; SH3.1.
KW Myosin; Actin-binding; ATP-binding; SH3 domain; Multigene family;
KW Chemotaxis; Phosphorylation.
FT DOMAIN 1 694 MYOSIN HEAD-LIKE.
FT DOMAIN 695 921 TAIL HOMOLOG REGION 1 (TH.1).
FT DOMAIN 922 1052 GLY/PRO/ALA-RICH (TH.2).
FT DOMAIN 1053 1111 SH3.
FT NE_BIND 102 109 ATP (POTENTIAL).
FT DOMAIN 547 627 ACTIN-BINDING.
FT DOMAIN 951 1015 ASN-RICH.
FT MOD_RES 332 332 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 1111 AA; 124313 MW; CD6366F08DC5642F CRC64;

```

```

Query Match 31.4%; Score 1692.5; DB 1; Length 1111;
Best Local Similarity 36.4%; Pred. No. 9..9e-100;
Matches 384; Conservative 172; Mismatches 312; Indels 187; Gaps 21;

```

```

QY 28 GVQDFVLENFTESEAFTELRERRENLITYTIGPVLSVNPYRDLQITSRQMERKRG 87
DB 10 GTDDLVLWPK-VSEDEICENLKRKRYMDFTYINGPVLISVNFRLNNSGPDETEAYR 68
QY 88 VSEFEVPHLEAVADVYRRLATERRDQAVMIGESGAGTEATKRLQFYAETCPAPER 147
DB 69 KHAQEVPRHYQLAESRYRAMKNDQEQVCIIGESAGCTEAKLIMGVSAISGTEK 128
QY 148 GCAVRDRLLQSNVLEAFGNAKTLRNDNSRFCKYMDVDFEGAPVGHILSYLLEKSR 207
DB 129 VEXYKHAIVLESNPLLEFGNAKTLRNNNSRFCKYFEIQDKAGDPVGKIVYLLKSR 188
QY 208 VVHQNIGERFHFYVYOLGEGEETLRRLGLENNPOSYLYVGCACAKVSSINDKPMV 267
DB 189 VVYQNGERFHFYVYOLGEGEETLRRLGLENNPOSYLYVGCACAKVSSINDKPMV 247
QY 268 MKRALSVIOETEDVEDDLSTAVSLHLCNHFADSDNSNAQVTTEKQLYETRLLEGV 327
DB 248 VROAMDITIGLTAQEQSDIIRIVACVLIHIGNYFTEDDKMAALYIDNALSLASMLCIDS 307
QY 328 TTIREALTHRIKTIAGK-----EELSPLENEQAAVADALAKAVYKFTFWLVKTN 379
DB 308 ATLQNALIFRVNTGAGAGAGNRSTYVNPQVNEQANGTFDALARTIYDRMFSLVEKVA 367
QY 380 RSLASDAESPWSRSTVYGLDIYGEFVQHNSEFQPCINVCNEKLOOLFIELTLKSRQ 439
DB 368 QSLSTY--YKSP---YONVIGILDIFFELFEKNGEFOFCINEVNEKLOOLFIELTLKSRQ 422
QY 440 EEEYAGSIAMEPVQYFNKNIICDLVEKFK-KGIISILDECLT--RPGATDTLTFLEKLE 495
DB 423 EEEVREGIKMEPIKRYFNQIVCOLLEGKSPGIFSLDIDICSTLHNSQSTDKCFLEKKA 482
QY 496 DTYKPRPHF--LTHKLADQKTRKSLDRGEFRLLHYAGEVYVSYVGLDKNNDLFENLKE 553
DB 483 GIYDGLHLMRGMT-----GAFALKHYAGEVYVSYVGLDKNNDLFENLKE 528
QY 554 TMCSSNPNIMAOCFDKSELS-DKKRPETVATQCFKMSLLDLVETLRSEKAYIRTCIRPNDA 612
DB 529 ALQCSAMPPLASLFNEDTSLQKKRPTTAGFKIKTSAGELMKALSCCTPHYIRCLIPNLT 588
QY 613 KQPGREDEVLIHQYKYLGLMENVRAAGFVRRKYEAFLQRYKSLCETPMMA---G 669
DB 589 KKAQDWENSRVKNHQVYQLLLEIVRVRAGFVRRKYEAFLQRYKSLCETPMMA---G 648
QY 670 RQDDGVAVLVRLGYKPEYKMGKRTKIFTRPKTLFATDSLEVRQSLATYIOAAMGEF 729

```

```

DB 649 DATEGCKTIFQDMNLEAGWQLOKTKVFIHNPETVLEBALKDKPDCTAKIQAKR-- 706
QY 730 HMRQKFLRVKRSALCIQSWMRGTLGRRAKAKRRMAQTIRRLRGLTRHSPRCPENAPF 789
DB 707 NMRKAK----- 711
QY 790 LDHVRSFLLNLRLQPLRVNLDTSWPTPPALREASELLRELCMKRMWYKCSISPEKK 849
DB 712 -----K 712
QY 850 QOLQKAVASEIFKGRKNDYPOGPRFLSTRLGTTEE---ISPRVLSLGSPEIQYAVPV 906
DB 713 HSLEQRAQLAHMFKDKKERQNSIDKRFSDYIDFENQGLQCAMONAKKKEVYRADVY 772
QY 907 VKIDRKGKPRPQOLLTTPSAVYIVEDAKYK-----QRIDYANLTSIVSSLSLSLF 958
DB 773 IKIDRRA-KQKNYENVLTQALYFVERKSILKKVLTILRVGLRREIKGSISTLSDNYI 831
QY 959 VLVHQRD-----KKQGDVLOSdhVIEFTLTALS---ADRVNNININQSGITTFAGPG 1011
DB 832 VFHLPEHDVYIENDKTEIILVLEYFKAGGSLWQFSDRIN-----YTLKKGQ 883
QY 1012 RDGIIDFTSGSE--LLITKAKNGHLAVAPRLNS 1043
DB 884 KE--ISFOKSEQCPILVVKKGKGLGTIASGLPS 916

```

```

RESULT 15
ID MYSD_DICDI STANDARD; PRT; 1113 AA.
AC P34109;
DT 01-FEB-1994 (rel. 28, Created)
DT 01-FEB-1994 (rel. 28, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Myosin ID heavy chain.
GN MYOD OR DMTD.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Dictyostelida; Dictyostelium.
OX NCBI_TaxId=4689;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 604-610; 733-742 AND 914-928.
RC STRAIN=AX3;
RA MEDLINE=93315475; PubMed=8325874;
RX Jung G., Fukui Y., Martin B., Hammer J.A. III;
RT "Sequence, expression pattern, intracellular localization, and
RT targeted disruption of the Dictyostelium myosin ID heavy chain
RT isoform."
RL J. Biol. Chem. 268:14981-14990(1993).
CC -!- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
CC ACTIVITY THAT IS ACTIVATED BY ACTIN. MYOSIN ID MAY HAVE A ROLE IN
CC CHEMOTAXIS AND AGGREGATION. IT COULD SERVE TO STABILIZE AND EVEN
CC RETRACT CORTICAL STRUCTURES, SUCH AS PSEUDOPODS AND LABELLOPODS.
CC -!- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED. DIMER OF A HEAVY
CC AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.
CC -!- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION JUST BENEATH THE
CC PLASMA MEMBRANE IN THE ANTERIOR PSEUDOPOD AT THE LEADING EDGE OF
CC THE CELL.
CC -!- DOMAIN: TH.1 BINDS DIRECTLY TO ANIONIC PHOSPHOLIPID MEMBRANES;
CC MYOSIN I COULD THEREFORE MOVE ACTIN RELATIVE TO MEMBRANES AND
CC VICE VERSA. TH.2 AND SH3 BIND TIGHTLY TO F-ACTIN; THIS TOGETHER
CC WITH THE NUCLEOTIDE-SENSITIVE SITE IN THE HEAD, ALLOWS SINGLE
CC MOLECULES OF MYOSIN I TO CROSS-LINK ACTIN FILAMENTS.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

```

DR EMBL: L16509; -: NOT\_ANNOTATED\_CDS.  
 DR PIR: A47106; A47106.  
 DR HSSP: P29354; 1GRI.  
 DR DictyDb: DD01048; myod.  
 DR InterPro: IPR001452; SH3.  
 DR InterPro: IPR001609; myosin\_head.  
 DR Pfam: PF00018; SH3; 1.  
 DR Pfam: PF00063; myosin\_head; 1.  
 DR PRINTS: PR00193; MYOSINHEAVY.  
 DR PRINTS: PR00452; SH3DOMAIN.  
 DR ProDom: PD000066; SH3; 1.  
 DR ProDom: PD000355; myosin\_head; 1.  
 DR SMART: SM00242; MYSC; 1.  
 DR SMART: SM00326; SH3; 1.  
 DR PROSITE: PS50002; SH3; 1.  
 KM Myosin: Actin-binding; ATP-binding; Chemotaxis; SH3 domain;  
 KM Multigene family.  
 FT DOMAIN 1 680 MYOSIN HEAD-LIKE.  
 FT DOMAIN 681 961 TAIL HOMOLOG REGION 1 (TH.1).  
 FT DOMAIN 962 1021 SH3.  
 FT DOMAIN 1022 1110 GLY/PRO/ALA-RICH (TH.2).  
 FT NP\_BIND 101 108 ATP (BY SIMILARITY).  
 FT DOMAIN 1016 1113 ALA/GLY/PRO-RICH.  
 SQ SEQUENCE 1113 AA; 124637 MW; 07FEB6C7FA7C2A08 CRC64;

Query Match 29.7%; Score 1601; DB 1; Length 1113;  
 Best Local Similarity 35.8%; Pred. No. 6,5e-94;  
 Matches 372; Conservative 177; Mismatches 301; Indels 188; Gaps 26;

QY 28 GVDPFVLENTSEAFATLNLRRFRRLNLIYTYIGPVLSVNPYRDIQ-IYSRQHMERIR 86  
 DB 8 GVDDMVMLSKIAND-SILDNLKRRYGGDVITYTYIGNVLISVNPFRKQIKNLYSERNLLEYR 66  
 QY 87 GVSFYEVPPLFAADVYVYRALRTERRDOAVMISGESGAGTEATKRLDQVAFATCPAPE 146  
 DB 67 GKFRYELPPHAYAVADDMYMSMAEGOSQCYIISGESGAGTEAKLIMOYIAVSGKA 126  
 QY 147 RCGAVRDRLLOSNFVLEAFNAKTLRNDNSRFGKMDVOFDFKAPVGHILSYLEKS 206  
 DB 127 DVSRYKVILLESNPLLEAFGAKTLRNNNSRFGKYMEOVNGIDEGGRTVYLYLEKS 186  
 QY 207 RVVHQNNGERNFHFVYOLLBEGEETLRRLGLERNPOSYLYLVKGCCKAVSSINDKSDWK 266  
 DB 187 RVVYQTKGERNFHFYOLLGSGANOQLKSELRLD-TPDKFNVLASGCVYVDGVDGSEFQ 245  
 QY 267 VMKALVIDFTEDVEDLSTIVASVHLGNIHFADEDSNAQVTTENQ--LKYTLRLG 324  
 DB 246 DVCAMKRVIGLTDEQKEVFLVAIILYLVGNVGRKNNAKDEAIDQOSKALENFAELMQ 305  
 QY 325 VEGTTLREALTHRKI-----IAKGEELLSPNLLEQAAVARDALAKAVYSRTFTWLVRK 377  
 DB 306 TDVSSCEKALCFRTISGTQGRSARVSTYACPONSEGAYISRDALAKALYSRLFDWIYGR 365  
 QY 378 INNSLASKDAESPWSRSTVYGLDIYGFVYQHNSEFQFCINYCNEKLOOLFTELTKS 437  
 DB 366 VNSALGYKONS-----OSLMGIDYGFELFEKNGFEQWVINYNERLOQIFTELTKT 420  
 QY 438 EOEYEAEGIAMEPVQYFNKKIICDLVE-EKFKGIISLDECLRPGEARPLTFLEKLED 496  
 DB 421 EOEYFNEGLOMEIODYFNKKICDLIESKPAIGILTLDDVCNP-KGDQKFLDLRKE 479  
 QY 497 TVKPHPLFTLHKLADQKTRKSLDRGEFFRLHYAGEVYTSVTFGLDKNNDLFLRNLKETMC 556  
 DB 480 SFSHNAHF-----QSAADS--SSSFTIKHYADVEYCAEGFYDKKNDLFLNDVELAA 530  
 QY 557 SSMPPIAQCQFDSKL-SDKRRPETAATQFKMSLLQVLEILRSKEPAYIRCIKPNDAKQP 615  
 DB 531 CTYSKLIPQLPEINCEKDKRKPTTAGFKIKESIGALVKALISACTPHYIRCIKPNKRA 590  
 QY 616 GRPDEVILRHQVYKTLGLMENL---RVRRAGFAIRKRYEAFLORYKSLCPETPWMAAGRP 671  
 DB 591 NDFDTSLVMHQVYKTLGLEENVRIFGRIRAGAYARQYTDKFFRYRVCCCKETWPNWTGCF 650

QY 672 QDGVAVLVRHLGKYPPEEYKMGRTKIFIRFPKTLFATEDSLSEVRROSLATKIQAAAMRGFHW 731  
 DB 651 ESGVETILKSWDLEPKQYSGKTKIFIRAPETVFNLE--ELRERKV-----FTY 697  
 QY 732 RQKFLRVKRSALCIQSWMRGTLGRKRAAKRMAOTIRRLIRGFIILHNSPCPENAFELD 791  
 DB 698 RNNKLQR-----FFLR 707  
 QY 792 HVRSFLLNLRLQPLRVNLDTSWPTPPALREASELRELCKNNMVKYCRSISPEMKQ 851  
 DB 708 FTLMSTYSIO-----KGAADSKSN 728  
 QY 852 LQKAVASE-IFGKKDNDYPOSVPRLEIFSTRGTETISPRYLOSGSEPIQYAVPVKYD 910  
 DB 729 KERRRLIERPYQGDYINRENF-----ELKD-IVKKNNGEKIMFTHAVVKYD 775  
 QY 911 RKGYPRPROLLTPSAVIV-----EDAK-----VKQIDVLANLTGISVSSISD 955  
 DB 776 RRSRCQR-RVLLSDTAIFYFATEKNKDKEDRKRPMIYVOKRRLLAGITSVBLSKLSD 834  
 QY 956 SLFVLHYQREDNKKQGVVLQSDHVIETLTK-----TALSADRVNNININ-OGSITFA-- 1007  
 DB 835 GFVVLKMANE-----HDOIPECRKTEPLGLIKAYKKTGTLRINYNNSIGVAIK 883  
 QY 1008 ---GPGRDGIIDFTSG 1021  
 DB 884 ASKQGGKGERILLFEKG 901

Search completed: December 9, 2002, 16:45:00  
 Job time: 32 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: December 9, 2002, 16:41:26 ; Search time 81 Seconds  
(without alignments)  
2655,718 Million cell updates/sec

Title: US-09-893-371-1

Perfect score: 5398  
Sequence: 1 MRYRASALGSDGVRTWESA.....LITKAKNGHLAVAPRLNSR 1044

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Database :

SPTREMBL\_21.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5398	100.0	1044	11 Q9ERB6	Q9ERB6 mus musculu
2	5261	97.5	1028	11 Q63355	Q63355 ratius norv
3	5128	95.0	1028	6 Q27966	Q27966 bos taurus
4	5111	94.7	1028	6 Q28138	Q28138 bos taurus
5	4371	81.0	1028	13 Q92002	Q92002 rana catesb
6	2629	48.7	1035	5 Q8R008	Q8R008 drosophila
7	1865	34.5	1006	11 Q63357	Q63357 ratius norv
8	1756	33.5	1017	5 Q27328	Q27328 caenorhabd
9	1719.5	31.9	1099	13 Q90748	Q90748 gallus gall
10	1715	31.8	1109	4 Q12965	Q12965 homo sapien
11	1714.5	31.8	1098	4 Q8WMN7	Q8WMN7 homo sapien
12	1712.5	31.7	1107	11 Q63356	Q63356 rattus norv
13	1695	31.4	1053	5 Q9NGI7	Q9NGI7 strongyloce
14	1573.5	29.1	1186	5 Q61080	Q61080 acanthameob
15	1564	29.0	1100	5 Q19901	Q19901 caenorhabd
16	1545	28.6	1249	3 Q00647	Q00647 emericella

17	1509.5	28.0	1217	3 Q9Y728	Q9Y728 schizosacch
18	1506	27.9	1049	5 Q00851	Q00851 entamoeba h
19	1504	27.9	633	4 Q96R16	Q96R16 homo sapien
20	1501	27.8	633	4 Q96R15	Q96R15 homo sapien
21	1473	27.3	1215	5 Q77202	Q77202 acanthameob
22	1405	26.0	1036	5 Q17382	Q17382 caenorhabd
23	1382.5	25.6	1446	10 Q9SVT9	Q9SVT9 arabidopsis
24	1362	25.2	2167	5 Q9V3Z6	Q9V3Z6 drosophila
25	1361	25.2	2357	5 Q9U1M8	Q9U1M8 dictyosteli
26	1355	25.1	1505	10 Q9LKB9	Q9LKB9 arabidopsis
27	1352	25.0	1515	10 Q39158	Q39158 arabidopsis
28	1346	24.9	839	10 Q9M0G3	Q9M0G3 arabidopsis
29	1336.5	24.8	2179	13 Q9M0G8	Q9M0G8 brachydanio
30	1336	24.7	1528	10 Q24518	Q24518 heliantus
31	1335	24.7	2121	5 Q9NH54	Q9NH54 drosophila
32	1333.5	24.7	2179	13 Q9DGG9	Q9DGG9 brachydanio
33	1331	24.7	1477	10 Q64491	Q64491 arabidopsis
34	1331	24.7	2062	11 Q9JUT5	Q9JUT5 mus musculu
35	1330.5	24.6	2113	11 Q99M26	Q99M26 mus musculu
36	1328	24.6	1547	10 Q8W312	Q8W312 oryza sativ
37	1328	24.6	1611	10 Q8RYP8	Q8RYP8 arabidopsis
38	1328	24.6	2129	5 Q9VLT3	Q9VLT3 drosophila
39	1326	24.6	1260	10 Q24516	Q24516 heliantus
40	1314	24.3	1242	10 Q9M2K0	Q9M2K0 arabidopsis
41	1313.5	24.3	1511	10 Q9M5A6	Q9M5A6 valisneria
42	1312	24.3	1490	10 Q9SKR0	Q9SKR0 arabidopsis
43	1311.5	24.3	1520	10 Q39160	Q39160 arabidopsis
44	1311	24.3	1515	10 Q9ATB7	Q9ATB7 petroselinu
45	1310.5	24.3	1529	10 Q9XEL4	Q9XEL4 zea mays (m

## ALIGNMENTS

```
RESULT 1
ID Q9ERB6 PRELIMINARY; PRT; 1044 AA.
AC Q9ERB6;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Nuclear myosin I beta.
GN MYOIC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Pestic-Dragovich L., Stojiljkovic L., Phillimonenko A.A., Nowak G.,
RA Ke Y., Settlage R.E., Shabanowitz J., Hunt D.F., Hozak P.,
RA de Lanerolle P.;
RT "A Myosin I Isoform in the Nucleus.";
RL Science 0:0-0(2000).
DR EMBL: AY007255; AAG02570.1; -.
DR HSSP: P08799; 1MND.
DR MGD: MGI:106612; Myo1c.
DR InterPro: IPR000048; IQ_region.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00612; IQ; 3.
DR Pfam: PF00063; myosin_head; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PDD00355; myosin_head; 1.
DR SMART: SM00015; IQ; 2.
DR SMART: SM00242; MYSC; 1.
DR SMART: SM00242; MYSC; 1.
SQ SEQUENCE 1044 AA; 119876 MW; CDAB0F699D041C9 CRC64;

Query Match 100.0%; Score 5398; DB 11; Length 1044;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1044; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
|||||1 MRYRASALGSDGVRTWESALTARDVGVDVFLLENFTSEAAFTLNLRPRRENTLYTY 60
|||||
```

```
Db 1 MRYASALGSDGVAVTMESALTARDRVGODPVLLENFTSEAFIENLRKRRENLITYT 60
QY 61 IGPVLVSNPRDLOIYSRQHEBRYRGVSEYVPHLFVAVDTYRALTERRRQAWMIS 120
Db 61 IGPVLVSNPRDLOIYSRQHEBRYRGVSEYVPHLFVAVDTYRALTERRRQAWMIS 120
QY 121 GEGSAGKTEATKRLIQEYAEETCPAPERGAVERLLOSNPYLEAFGNAKTLRDNSSRF 180
Db 121 GEGSAGKTEATKRLIQEYAEETCPAPERGAVERLLOSNPYLEAFGNAKTLRDNSSRF 180
QY 181 KYMVOVDFKAPVGVGHILSLYLEKSRVAVQNHGRNPHVYQULEGGEETLRRLGLER 240
Db 181 KYMVOVDFKAPVGVGHILSLYLEKSRVAVQNHGRNPHVYQULEGGEETLRRLGLER 240
QY 241 NPOSYLYLVKGQCAKASSINDKSDMKVMRKALSYIDFTEDEVEDLSTIVASVHLGINHF 300
Db 241 NPOSYLYLVKGQCAKASSINDKSDMKVMRKALSYIDFTEDEVEDLSTIVASVHLGINHF 300
QY 301 AAEDSNAQVTTENOLKYTLRLGVEGTTLRALTLHRKIIAKGEBLLSPNLEQAAYARD 360
Db 301 AAEDSNAQVTTENOLKYTLRLGVEGTTLRALTLHRKIIAKGEBLLSPNLEQAAYARD 360
QY 361 ALKAVYSRFTWLVKIRKINSLSKDAESPSWRSTYVLGLDITYGEFVQHNSEFCIN 420
Db 361 ALKAVYSRFTWLVKIRKINSLSKDAESPSWRSTYVLGLDITYGEFVQHNSEFCIN 420
QY 421 YCNEKLOQLELTLELTKSEOEYEAEIGIAMEPVQYFNKKIICDLVEEKFCTIILDEEC 480
Db 421 YCNEKLOQLELTLELTKSEOEYEAEIGIAMEPVQYFNKKIICDLVEEKFCTIILDEEC 480
QY 481 RPEGATDLTLEKLEDTYVKKPHFLTHKLADOKTRKSLDNGEERLLHYAGEVITYSVTGFL 540
Db 481 RPEGATDLTLEKLEDTYVKKPHFLTHKLADOKTRKSLDNGEERLLHYAGEVITYSVTGFL 540
QY 541 DKNNDLFLRLKJETMCSMNPIMACDFKSELSDKKRPETVATQFKMSLQVLEILRSKE 600
Db 541 DKNNDLFLRLKJETMCSMNPIMACDFKSELSDKKRPETVATQFKMSLQVLEILRSKE 600
QY 601 PAVIRICIRPDARQGFHDEVLIRHOVKYLGLENLVRKACFAYRRKYAFAFIQRRKSLC 660
Db 601 PAVIRICIRPDARQGFHDEVLIRHOVKYLGLENLVRKACFAYRRKYAFAFIQRRKSLC 660
QY 661 PETWPMWAGRPQGVAVLVNHLGKREYKMGRTKIFIRPKTLFATEDSLEVRQSLAT 720
Db 661 PETWPMWAGRPQGVAVLVNHLGKREYKMGRTKIFIRPKTLFATEDSLEVRQSLAT 720
QY 721 KIOAAMRGFHMROKFLRVKRSALCISQMWNGTIGRRKAARKMAAOTIRRLIGFILLRS 780
Db 721 KIOAAMRGFHMROKFLRVKRSALCISQMWNGTIGRRKAARKMAAOTIRRLIGFILLRS 780
QY 781 PRCPENAFLLDIHVASFLLNLRQLPRNVLDTSMPPTPPALREASELLRELCKNMVMKY 840
Db 781 PRCPENAFLLDIHVASFLLNLRQLPRNVLDTSMPPTPPALREASELLRELCKNMVMKY 840
QY 841 CRISISEMKGQLOOKAVASIFEGKADNTPQSVPLFISIRLCTEISPRVLOSICSEPT 900
Db 841 CRISISEMKGQLOOKAVASIFEGKADNTPQSVPLFISIRLCTEISPRVLOSICSEPT 900
QY 901 QYAVPVVKYDRKGYKRRPRLTLTPSAVYIVEDAKYKORIDYANLGISVSSLSDSLFLY 960
Db 901 QYAVPVVKYDRKGYKRRPRLTLTPSAVYIVEDAKYKORIDYANLGISVSSLSDSLFLY 960
QY 961 HVOERDNKQGVVLOSJHVIETLTALSAHVNININOGSITFAGGPGRGIDFTS 1020
Db 961 HVOERDNKQGVVLOSJHVIETLTALSAHVNININOGSITFAGGPGRGIDFTS 1020
QY 1021 GSELLITTKAKNGHLAVVAPRLNSR 1044
Db 1021 GSELLITTKAKNGHLAVVAPRLNSR 1044
```

RESULT 2  
063355

```
ID 063355 PRELIMINARY; PRT: 1028 AA.
AC 063355;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Myosin I heavy chain.
GN MYR 2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC SPRATIN-SPRAGUE-DAWLEY.
RA Ruppert C., Godel J., Reinhard J., Baehler M.;
RT "Myr-2 a novel class-I Myosin identified in rat brain.";
RL Submitted (Aug-1993) to the EMBL/Genbank/DBJ databases.
DR EMBL; X74800; CAA52807.1; -.
DR HSSP; P08799; 1MND.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00612; IQ; 3.
DR Pfam; PF00063; myosin_head; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 2.
DR SMART; SM00242; MYSC; 1.
SQ SEQUENCE 1028 AA; 118090 MW; 10EB314FD5996CB5 CRC64;

Query Match 97.5%; Score 5261; DB 11; Length 1028;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 1015; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 17 MESALTARDRVGODPVLLENFTSEAFIENLRKRRENLITYTIGPVLSVNPYRDLQI 76
Db 1 MESALTARDRVGODPVLLENFTSEAFIENLRKRRENLITYTIGPVLSVNPYRDLQI 76
QY 77 YSRQHEBRYRGVSEYVPHLFVAVDTYRALTERRRQAWMISGSGGKTEATKRLIQ 136
Db 77 YSRQHEBRYRGVSEYVPHLFVAVDTYRALTERRRQAWMISGSGGKTEATKRLIQ 136
QY 137 FYAETCPAPERGAVERLLOSNPYLEAFGNAKTLRDNSSRFKGYMVDQPFKAPYVG 196
Db 137 FYAETCPAPERGAVERLLOSNPYLEAFGNAKTLRDNSSRFKGYMVDQPFKAPYVG 196
QY 197 HILSTYLLKSRVAVQNHGRNPHVYQULEGGEETLRRLGLERNPQSTLYLVKGQCAK 256
Db 197 HILSTYLLKSRVAVQNHGRNPHVYQULEGGEETLRRLGLERNPQSTLYLVKGQCAK 256
QY 257 SSINDKSDMKVMRKALSYIDFTEDEVEDLSTIVASVHLGINHFAADESNQVTTENOL 316
Db 257 SSINDKSDMKVMRKALSYIDFTEDEVEDLSTIVASVHLGINHFAADESNQVTTENOL 316
QY 317 KYTLRLGVEGTTLRALTLHRKIIAKGEBLLSPNLEQAAYARDALAAVYSRFTWLV 376
Db 317 KYTLRLGVEGTTLRALTLHRKIIAKGEBLLSPNLEQAAYARDALAAVYSRFTWLV 376
QY 376 KINRSLASKDASPSPWRSTYVLGLDITYGEFVQHNSEFCINCNKLOQLFELTLK 436
Db 376 KINRSLASKDASPSPWRSTYVLGLDITYGEFVQHNSEFCINCNKLOQLFELTLK 436
QY 437 SGOEYEAEIGIAMEPVQYFNKKIICDLVEEKFCTIILDEECLEPGEATDLTLEKLED 496
Db 437 SGOEYEAEIGIAMEPVQYFNKKIICDLVEEKFCTIILDEECLEPGEATDLTLEKLED 496
QY 497 TYKPHPHFLTHKLADOKTRKSLDNGEERLLHYAGEVITYSVTGFLDKNNDLFLRLK 556
Db 497 TYKPHPHFLTHKLADOKTRKSLDNGEERLLHYAGEVITYSVTGFLDKNNDLFLRLK 556
QY 557 SSMNPIMACDFKSELSDKKRPETVATQFKMSLQVLEILRSKEPARYIRCIKPNDAKPG 616
Db 557 SSMNPIMACDFKSELSDKKRPETVATQFKMSLQVLEILRSKEPARYIRCIKPNDAKPG 616
```

```
QY 617 REDEVLIIRHGVKYLGLMELRVRRAGFAVRKYEAFLOKYSLSCEPTWPMAGRPDQGYA 676
Db 601 RDEVELIIRHGVKYLGLMELRVRRAGFAVRKYEAFLOKYSLSCEPTWPMAGRPDQGYA 660
QY 677 VLVRLHGYKPEEYKMGRTKIFIRFPKTLFATPDSLEVRROSLATKIOAAMRGFHHROKFL 736
Db 661 VLVRLHGYKPEEYKMGRTKIFIRFPKTLFATPDSLEVRROSLATKIOAAMRGFHHROKFL 720
QY 737 RYKRSALICQSWMRGTLGRKRAKRAKMAOQTRILRGFILRHSPRCPENAFLDHWARS 796
Db 721 RYKRSALICQSWMRGTLGRKRAKRAKMAOQTRILRGFILRHSPRCPENAFLDHWARS 780
QY 797 FLNLNRROLPRNVLDTSWPTPPALREASELLRELCKMMVMKCYCRSISPEMKOOLQOKA 856
Db 781 FLNLNRROLPRNVLDTSWPTPPALREASELLRELCKMMVMKCYCRSISPEMKOOLQOKA 840
QY 857 VASELFFKGGKNDYPOSVPRLFISTRLGTEELSPRYLOSLGSEPIQYAVVYVYDKRGYKP 916
Db 841 VASELFFKGGKNDYPOSVPRLFISTRLGTEELSPRYLOSLGSEPIQYAVVYVYDKRGYKP 900
QY 917 RPRQLLTPSAVVIYEDAKVKQKIDYANLTGIVSVSLSDSLFVLHVOREDNKOKGDVYLQ 976
Db 901 RSRQLLTPSAVVIYEDAKVKQKIDYANLTGIVSVSLSDSLFVLHVOREDNKOKGDVYLQ 960
QY 977 SDHVETLTKTALSADRVNNININOGSITTFAGGPGHDGIIDFTSGSELLITAKNGHLAV 1036
Db 961 SDHVETLTKTALSADRVNNININOGSITTFAGGPGHDGIIDFTSGSELLITAKNGHLAV 1020
QY 1037 VAPRLNSR 1044
Db 1021 VAPRLNSR 1028

RESULT 3
QY 027966 PRELIMINARY: PRT: 1028 AA.
AC 027966
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Myosin I.
OS Bos taurus (bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N. A.
RC TISSUE=ADRENAL GLAND;
RX MEDLINE=9414808; PubMed=8313976;
RA Zhu T., Ikebe M.;
RT "Cloning of myosin I from bovine adrenal gland.";
RL FEBS Lett. 339:31-36(1994).
DR EMBL: U03420; AAA17565.1; -.
DR HSSP: P08799; 1MD.
DR InterPro: IPR000048; 1Q_region.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00612; 1Q; 2.
DR PRINTS: PR00193; myosin_head.1.
DR PRODOM: PD000355; myosin_head.1.
DR SMART: SM00015; 1Q; 2.
DR SMART: SM00242; MYSC; 1.
SQ SEQUENCE 1028 AA; 118020 MW; 5CFEPE28CFZA5D7 CRC64;
```

Query Match 95.0%; Score 5128; DB 6; Length 1028;  
Best Local Similarity 95.6%; Pred. No. 0;  
Matches 983; Conservative 25; Mismatches 20; Indels 0; Gaps 0;

```
QY 17 MESALTDARDVGVOPVLENTFTSEAFIENLRPRRENLYTYIGPVVSNPYRDLQI 76
Db 1 MESALTARDVGVOPVLENTFTSEAFIENLRPRRENLYTYIGPVVSNPYRDLQI 60
```

```
QY 77 YSRQHMERYRGVSFEVPRPHLFAVADTVYRALRTERDDAVMISGSGAKTEATKRLQ 136
Db 61 YSRQHMERYRGVSFEVPRPHLFAVADTVYRALRTERDDAVMISGSGAKTEATKRLQ 120
QY 137 FYAETCPAPERGGAARDRLQSNPLYEAFGNKTLENDSSRFKGKMDYQDFPKGAPVG 196
Db 121 FYAETCPAPERGGAARDRLQSNPLYEAFGNKTLENDSSRFKGKMDYQDFPKGAPVG 180
QY 197 HILSYLLEKSRVYHONHGRNFHVEYQLLEGGEEETLRILGLEERNPOSYLYLVKGCARV 256
Db 181 HILSYLLEKSRVYHONHGRNFHVEYQLLEGGEEETLRILGLEERNPOSYLYLVKGCARV 240
QY 257 SSINDKSDMKVNRKALSVITDFTDEVEDLLSTVASYLHICNTHPAADESSNAQVTTENOL 316
Db 241 SSINDKSDMKVNRKALSVITDFTDEVEDLLSTVASYLHICNTHPAADESSNAQVTTENOL 300
QY 317 KYTLRLGVEGTTLRPAHLRHKTIKAGEELLSPINLEQAAVARDALAKAVYSTFTWILVR 376
Db 301 KYTLRLGVEGTTLRPAHLRHKTIKAGEELLSPINLEQAAVARDALAKAVYSTFTWILVA 360
QY 377 KINRSILASKDAESPWSRSTTVGLLDIYGFVEYQHNSEFQFCINYNENKLOLFIELTLK 436
Db 361 KINRSILASKDAESPWSRSTTVGLLDIYGFVEYQHNSEFQFCINYNENKLOLFIELTLK 420
QY 437 SQDEYEAEAGIAMEPVQYVNNKICDLVEEKFGIISILDEDECLRGEATDLFLEKLED 496
Db 421 SQDEYEAEAGIAMEPVQYVNNKICDLVEEKFGIISILDEDECLRGEATDLFLEKLED 480
QY 497 TVKPRPHPLTHKLADOKTKKSLDRGEFLLHYAGEVTSVGTGLKNNDLPRNLKETMC 556
Db 481 TVKPRPHPLTHKLADOKTKKSLDRGEFLLHYAGEVTSVGTGLKNNDLPRNLKETMC 540
QY 557 SSMPNIMAOCPDKSLSLSDKKRPETVATQFKMSLLQLEVLIRKSEPAVYIRICRPNDAKOPG 616
Db 541 SSMPNIMAOCPDKSLSLSDKKRPETVATQFKMSLLQLEVLIRKSEPAVYIRICRPNDAKOPG 600
QY 617 REDEVLIIRHGVKYLGLMELRVRRAGFAVRKYEAFLOKYSLSCEPTWPMAGRPDQGYA 676
Db 601 RDEVELIIRHGVKYLGLMELRVRRAGFAVRKYEAFLOKYSLSCEPTWPMAGRPDQGYA 660
QY 677 VLVRLHGYKPEEYKMGRTKIFIRFPKTLFATPDSLEVRROSLATKIOAAMRGFHHROKFL 736
Db 661 VLVRLHGYKPEEYKMGRTKIFIRFPKTLFATPDSLEVRROSLATKIOAAMRGFHHROKFL 720
QY 737 RYKRSALICQSWMRGTLGRKRAKRAKMAOQTRILRGFILRHSPRCPENAFLDHWARS 796
Db 721 RYKRSALICQSWMRGTLGRKRAKRAKMAOQTRILRGFILRHSPRCPENAFLDHWARS 780
QY 797 FLNLNRROLPRNVLDTSWPTPPALREASELLRELCKMMVMKCYCRSISPEMKOOLQOKA 856
Db 781 FLNLNRROLPRNVLDTSWPTPPALREASELLRELCKMMVMKCYCRSISPEMKOOLQOKA 840
QY 857 VASELFFKGGKNDYPOSVPRLFISTRLGTEELSPRYLOSLGSEPIQYAVVYVYDKRGYKP 916
Db 841 VASELFFKGGKNDYPOSVPRLFISTRLGTEELSPRYLOSLGSEPIQYAVVYVYDKRGYKP 900
QY 917 RPRQLLTPSAVVIYEDAKVKQKIDYANLTGIVSVSLSDSLFVLHVOREDNKOKGDVYLQ 976
Db 901 RSRQLLTPSAVVIYEDAKVKQKIDYANLTGIVSVSLSDSLFVLHVOREDNKOKGDVYLQ 960
QY 977 SDHVETLTKTALSADRVNNININOGSITTFAGGPGHDGIIDFTSGSELLITAKNGHLAV 1036
Db 961 SDHVETLTKTALSADRVNNININOGSITTFAGGPGHDGIIDFTSGSELLITAKNGHLAV 1020
QY 1037 VAPRLNSR 1044
Db 1021 VAPRLNSR 1028
```

RESULT 4  
QY 028138 PRELIMINARY: PRT: 1028 AA.  
AC 028138

DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Myosin IB.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 NC NCB1\_TaxID=9913;  
 RN NCB1  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RX MEDLINE=94294379; PubMed=8022785;  
 RA Reizes O., Barylko B., Li C., Sudhof T.C., Albenist J.P.;  
 RT "Domain Structure of a mammalian myosin I-beta."  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:6349-6353(1994).  
 DR EMBL; Z22852; CAA80476.1; -  
 DR HSSP; P08799; 1MND.  
 DR InterPro; IPR000048; IQ\_region.  
 DR InterPro; IPR001609; myosin\_head.  
 DR Pfam; PF00612; IQ; 2.  
 DR Pfam; PF00063; myosin\_head; 1.  
 DR PRINTS; PR00193; MYOSINHEAVY.  
 DR Prodom; PD000355; myosin\_head; 1.  
 DR SMART; SM00015; IQ; 2.  
 DR SMART; SM00242; MYSC; 1.  
 DR SMART; SM00242; MYSC; 1.  
 SQ SEQUENCE 1028 AA; 118241 MW; 95E7E806CC86B2CD CRC64;

Query Match 94.7%; Score 5111; DB 6; Length 1028;  
 Best Local Similarity 95.3%; Pred. No. 0;  
 Matches 980; Conservative 27; Mismatches 21; Indels 0; Gaps 0;

QY 17 MESALTRDPRGVQDVLENTFTSEAFTEENLRRENTIYTGVLVSNPYRDLQ 76  
 DB 1 MESALTRDPRGVQDVLENTFTSEAFTEENLRRENTIYTGVLVSNPYRDLQ 60  
 QY 77 YSRQHMERKRGVSFEVPHLEAVADTVYRALFTRERDQAVMTSGSAGTEATRLQ 136  
 DB 61 YSRQHMERKRGVSFEVPHLEAVADTVYRALFTRERDQAVMTSGSAGTEATRLQ 120  
 QY 137 FYAETCPAPRGAVRDLLOSMPVLEAFGMATLNDNSRRGKMDVQDFDKGAPVG 196  
 DB 121 FYAETCPAPRGAVRDLLOSMPVLEAFGMATLNDNSRRGKMDVQDFDKGAPVG 180  
 QY 197 HILSYLLEKSRVYHQNNGERNHFVQLLEGCEETLRRLGLEERNQSYLYLVKGCCKAY 256  
 DB 181 HILSYLLEKSRVYHQNNGERNHFVQLLEGCEETLRRLGLEERNQSYLYLVKGCCKAY 240  
 QY 257 SSINDKSMKVMRKALSVDFTEDEVEDLSTIVASVLHNGNHFADEDSNAOVTTENQ 316  
 DB 241 SSINDKSMKVMRKALSVDFTEDEVEDLSTIVASVLHNGNHFADEDSNAOVTTENQ 300  
 QY 317 KYTLRLGEGTTLRALHLRKIIAGBELSLNLEQAAVADALAKAVYSTFTWLVK 376  
 DB 301 KYTLRLGEGTTLRALHLRKIIAGBELSLNLEQAAVADALAKAVYSTFTWLVK 360  
 QY 377 KINRSKASDAESPWSRSTTVGLLDIYGEVEYQHNSPFQFCINYNENKLOQFIFELTK 436  
 DB 361 KINRSKASDAESPWSRSTTVGLLDIYGEVEYQHNSPFQFCINYNENKLOQFIFELTK 420  
 QY 437 SQGEVEAGIAMEPVQYFNKKICDLVEEKFQGISILDEELRPGCATDLTFLEKLD 496  
 DB 421 SQGEVEAGIAMEPVQYFNKKICDLVEEKFQGISILDEELRPGCATDLTFLEKLD 480  
 QY 497 TVKPHPHPLTHKLADOKTRKSLDRGEFRLIHAGEVYTVTGFLDKNNDLFRNLKETWC 556  
 DB 481 TVKPHPHPLTHKLADOKTRKSLDRGEFRLIHAGEVYTVTGFLDKNNDLFRNLKETWC 540  
 QY 557 SSMNPIMACQFDSKSLSDKKRPETVATQFKMSLDQVELLRSEPAVYICIRPNDAKQKG 616  
 DB 541 SSMNPIMACQFDSKSLSDKKRPETVATQFKMSLDQVELLRSEPAVYICIRPNDAKQKG 600  
 QY 617 RFDEVILIRHQVYKLGIMENLRVRAGFAVRKYEAFLQRYKSLCPETWMMAGRPQDQVA 676

DB 601 RFDEVILIRHQVYKLGIMENLRVRAGFAVRKYEAFLQRYKSLCPETWMTGRPODQV 660  
 QY 677 VLVRHILGYKPEEKMGRTKFEIRPKTLFATETESLEVRRQSLATKIQOARGHMOKFL 736  
 DB 661 VLVRHILGYKPEEKMGRTKFEIRPKTLFATETESLEVRRQSLATKIQOARGHMOKFL 720  
 QY 737 RVKRSALCIQSWRGTLGRKRAKRAKMAQIIRLIRGLRISPRCPENAFPLDHWAS 796  
 DB 721 RLKRSALCIQSWRGTLGRKRAKRAKMAQIIRLIRGLRISPRCPENAFPLDHWAS 780  
 QY 797 FLNLRRQLPRNVLDTSWPPPALRBSLLELCKMKNWYKCSISEPWQOIQOKA 856  
 DB 781 FLNLRRQLPRNVLDTSWPPPALRBSLLELCKMKNWYKCSISEPWQOIQOKA 840  
 QY 857 VASEIFKGGKDNYPQSVPRFLFISTRGLTEBISPRVLOSISSEPIQYAVPVKDRKGYP 916  
 DB 841 VASEIFKGGKDNYPQSVPRFLFISTRGLTEBISPRVLOSISSEPIQYAVPVKDRKGYP 900  
 QY 917 RPRQLLTPSAVYVEDAKVKORIDYANLTGISVSSLSDSLFLVHQREDNKQGDVLD 976  
 DB 901 RSRQLLTPSAVYVEDAKVKORIDYANLTGISVSSLSDSLFLVHQREDNKQGDVLD 960  
 QY 977 SDHVIETLTKTALSADRVNNINIGSITFPAGPGSDGIIDFTSGSELLITKAKNGLAV 1036  
 DB 961 SDHVIETLTKTALSADRVNNINIGSITFPAGPGSDGIIDFTSGSELLITKAKNGLAV 1020  
 QY 1037 VAPRLNSR 1044  
 DB 1021 VAPRLNSR 1028

RESULT 5  
 ID 092002 PRELIMINARY; PRT: 1028 AA.  
 AC 092002;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Myosin I beta.  
 GN AMT BETA.  
 OS Rana catesbeiana (Bull frog).  
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Ranidae; Rana.  
 NC NCB1\_TaxID=8400;  
 RN NCB1  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RX MEDLINE=95083594; PubMed=7991542;  
 RA Metcalf A.B., Chelliah Y., Hudspeeth A.J.;  
 RT "Molecular cloning of a myosin I beta isoform that may mediate  
 adaptation by hair cells of the bullfrog's internal ear."  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:11821-11825(1994).  
 RN NCB1  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=KIDNEY;  
 RA Solc C.F., Derfler B.H., Duyk G.M., Corey D.P.;  
 RL Aud. Neurosci. 1:63-75(1994).  
 DR EMBL; U14549; AAA57192.1; -  
 DR EMBL; U14382; AAA65091.1; -  
 DR HSSP; P08799; 1MND.  
 DR InterPro; IPR000048; IQ\_region.  
 DR InterPro; IPR001609; myosin\_head.  
 DR Pfam; PF00612; IQ; 2.  
 DR Pfam; PF00063; myosin\_head; 1.  
 DR PRINTS; PR00193; MYOSINHEAVY.  
 DR Prodom; PD000355; myosin\_head; 1.  
 DR SMART; SM00015; IQ; 2.  
 DR SMART; SM00242; MYSC; 1.  
 SQ SEQUENCE 1028 AA; 118830 MW; D15F6F99B68A90AE CRC64;

Query Match 81.0%; Score 4371; DB 13; Length 1028;  
 Best Local Similarity 79.2%; Pred. No. 1.2e-292;

Matches 814; Conservative 112; Mismatches 102; Indels 0; Gaps 0;			
Qy	17	MESALTARDRVGVODFVLENTSEAFIENLRFRRENLIYTYIGVLSVNPYRDIQI	76
Db	1	MESALTARDRVGVODFVLENTSEAFIENLRFRRENLIYTYIGVLSVNPYRDIQI	60
Qy	77	YSROHMERVGVSYFVPPHLEAVADTVYRALRTRRQAVMISGESGAKTEATKRLQ	136
Db	61	YSKQOMERYRGVSFVSPHYAIADNSYRSLRTERKQOCILISGESGAKTEASKILQ	120
Qy	137	FYAECPAPREGAVRDLQSNPVLFAFGNAKTYLRNDSNRFKGYMVOEDFKAPVGG	196
Db	121	YAAVCPADVETVYKQDLQSNPVLFAFGNAKTYLRNDSNRFKGYMVOEDFKAPVGG	180
Qy	197	HLISLLEKSRVYHONHBERNFHVYOLLEGGEEETLRRLGERNPQSLYLVKQCAVY	256
Db	181	HLISLLEKSRVYHONHBERNFHVYOLLEGGEEETLRRLGERNPQSLYLVKQCAVY	240
Qy	257	SSINDSDMKVMKALSVIDFTEDEVEDLLSTIVASVHLGNIFHADEDSNAQVTTENQ	316
Db	241	SSINDSDMKVMKALSVIDFTEDEVEDLLSTIVASVHLGNIFHADEDSNAQVTTENQ	300
Qy	317	KYLTFLLOVEGTEALRTHKTIKAGEELSPVLNLEQAAVARDALAAVYRTTWLVR	376
Db	301	KYLTFLLOVEGTEALRTHKTIKAGEELSPVLNLEQAAVARDALAAVYRTTWLVR	360
Qy	377	KINRSLASKDAESPMSRSTTVYGLDIYGFVYFONHSEFOFCINCNKLOOLFETLTK	436
Db	361	KINRSLASKDAESPMSRSTTVYGLDIYGFVYFONHSEFOFCINCNKLOOLFETLTK	420
Qy	437	SEOEYEAEGIAMPEVOYFNNKIIDLVEEKKGIIISLDEBCLRPGEATDULFLEKLED	496
Db	421	SEOEYEAEGIAMPEVOYFNNKIIDLVEEKKGIIISLDEBCLRPGEATDULFLEKLED	480
Qy	497	TYKPRPHLTHLQAOQKRSLSDBREFLTHAGVYTSVYGLDKNDLLEFRNLKETHC	556
Db	481	TYKPRPHLTHLQAOQKRSLSDBREFLTHAGVYTSVYGLDKNDLLEFRNLKETHC	540
Qy	557	SSMNPIMOCFDPKSELSPKRPETAVATOFKMSLULVELTRKEPAYTRICIPNDAKOPG	616
Db	541	SSMNPIMOCFDPKSELSPKRPETAVATOFKMSLULVELTRKEPAYTRICIPNDAKOPG	600
Qy	617	REDEVLIHQVYKYLIGMENLVRRAAGFAYRRKYEAFLORYKSLCPETWPMAGRODGYA	676
Db	601	REDEVLIHQVYKYLIGMENLVRRAAGFAYRRKYEAFLORYKSLCPETWPMAGRODGYA	660
Qy	677	VLYRNLGKRPPEYKMGRTKIFTRPKTLFATPDSLEVRQSLATKIQAAKRGFHHKQKFL	736
Db	661	VLYRNLGKRPPEYKMGRTKIFTRPKTLFATPDSLEVRQSLATKIQAAKRGFHHKQKFL	720
Qy	737	RKRSALICIOSMWRGTLGRRAAKRKMAQOTIRRLIRGFIHSPRCPENAEFLDHVRAS	796
Db	721	RKRSALICIOSMWRGTLGRRAAKRKMAQOTIRRLIRGFIHSPRCPENAEFLDHVRAS	780
Qy	797	FLNLRLROLPRNVLDTSWTPPPALREASSELLRELCMKNMVKKYCRSISPEMKQOLOOKA	856
Db	781	FLNLRLROLPRNVLDTSWTPPPALREASSELLRELCMKNMVKKYCRSISPEMKQOLOOKA	840
Qy	857	VASELFGKCKNDYPOSVPRFLSTRLGTEISPRVLOSIGSEPIQYAVVYVYDKRGYR	916
Db	841	VASELFGKCKNDYPOSVPRFLSTRLGTEISPRVLOSIGSEPIQYAVVYVYDKRGYR	900
Qy	917	RPROLLTPSAVIVIEDAKVKORIDYANLTGVSLSLSLVLVHOREDNKQOGVYLO	976
Db	901	RPROLLTPSAVIVIEDAKVKORIDYANLTGVSLSLSLVLVHOREDNKQOGVYLO	960
Qy	977	SOHVLETLTKTALSADRVNNININOGSITFPAGPGRDGIIDFTSSSELLITRAKNGHLAV	1036
Db	961	SOHVLETLTKTALSADRVNNININOGSITFPAGPGRDGIIDFTSSSELLITRAKNGHLAV	1020
Qy	1037	VAPRLNSR 1044	
Db	1021	VAPRLNSR 1028	

RESULT 6			
ID	Q8T008	PRELIMINARY;	PRT; 1035 AA.
DT	01-JUN-2002 (TREMBlrel, 21, Created)		
DT	01-JUN-2002 (TREMBlrel, 21, Last sequence update)		
DT	01-JUN-2002 (TREMBlrel, 21, Last annotation update)		
DE	G04201p.		
GN	M061F.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephyroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BERKELEY;		
RA	Stapleton M., Brocstein P., Hong L., Abdayani A., Carlson J.,		
RA	Champe M., Chavez C., Dorsett V., Farfan D., Fise E., George R.,		
RA	Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,		
RA	Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,		
RA	Yu C., Lewis S.E., Rubin G.M., Celinker S.;		
RA	Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.		
DR	EMBL; AY069044; AAL39189.1;		
SR	SEQUENCE 1035 AA; 119023 MW; 6EE2C81E6456E63C CRC64;		
Query Match 48.7%; Score 2629; DB 5; Length 1035;			
Best Local Similarity 51.1%; Pred. No. 1.8e-172;			
Matches 533; Conservative 180; Mismatches 306; Indels 24; Gaps 8;			
Qy	8	IGSDGVRYMESALTARDRVGVODFVLENTSEAFIENLRFRRENLIYTYIGVLSV	67
Db	1	MASFNSQLMETGLHERDRAGVODFVLENTSEAFIENLRFRRENLIYTYIGVLSV	60
Qy	68	VNPYRDLQIYSROHMERVGVSYFVPPHLEAVADTVYRALRTRRQAVMISGESGACK	127
Db	61	VNPYRDLQIYSROHMERVGVSYFVPPHLEAVADTVYRALRTRRQAVMISGESGACK	120
Qy	128	TEATRLILOFVAETCPAPERGAVDRLQSNPVLFAFGNAKTYLRNDSNRFKGYMVOE	187
Db	121	TEATRLILOFVAETCPAPERGAVDRLQSNPVLFAFGNAKTYLRNDSNRFKGYMVOE	180
Qy	188	DKGAPVGGHILSYLLEKSRVYHONHBERNFHVYOLLEGGEEETLRRLGERNPQSLY	247
Db	181	DKGAPVGGHILSYLLEKSRVYHONHBERNFHVYOLLEGGEEETLRRLGERNPQSLY	240
Qy	248	LVKQCAKVSINDSDMKVMKALSVIDFTEDEVEDLLSTIVASVHLGNIFHADEDSN	307
Db	241	LVKQCAKVSINDSDMKVMKALSVIDFTEDEVEDLLSTIVASVHLGNIFHADEDSN	299
Qy	308	AOVTENOLKTYLRLGVCSTLRREALTHKTIKAGEELSPVLNLEQAAVARDALAAVY	367
Db	300	AKVNSRDVLVTAARLGVASLEALATHRTIDARDVVTSPLENELAYARDALAAVY	359
Qy	368	STFTFWLVKIRSLASKAESPMSRSTTVYGLDIYGFVYFONHSEFOFCINCNKLO	427
Db	360	STFTFWLVKIRSLASKAESPMSRSTTVYGLDIYGFVYFONHSEFOFCINCNKLO	416
Qy	428	OLFIETLTKSEOEYEAEGIAMPEVOYFNNKIIDLVEEKKGIIISLDEBCLRPGEATD	487
Db	417	OLFIETLTKSEOEYEAEGIAMPEVOYFNNKIIDLVEEKKGIIISLDEBCLRPGEATD	476
Qy	488	LTFLEKLEDTVPKPHFLTHKLADQKRSLSDBREFLTHAGVYTSVYGLDKNDL	547
Db	477	LTFLEKLEDTVPKPHFLTHKLADQKRSLSDBREFLTHAGVYTSVYGLDKNDL	536
Qy	548	FNLEKTMSSNMPIMOCFDPKSELSPKRPETAVATOFKMSLULVELTRKEPAYTRIC	607
Db	537	FNLEKTMSSNMPIMOCFDPKSELSPKRPETAVATOFKMSLULVELTRKEPAYTRIC	596
Qy	608	KPNDAKOPGRFDEVLIHQVYKYLIGMENLVRRAAGFAYRRKYEAFLORYKSLCPETWPM	667



```

Db 597 KPNLOQANVNDYLVHLQVYKYLGMENLRVRAGFAVRRTYELFLEKYSLSSTWPNY 656
Qy 668 --AARPODGVANVLRHLGCKRPEEKMGRTKIFIRPKTLFATEDSLEVRROSLATKIOAA 725
Db 657 KCPGPAAGVVOOLVKDGMDEKRYVGETKLFIRWPRTLFPTEDAYOEKKHEIAIIQAH 716
Qy 726 WRGFIWROKFLVRKSAICISWMRGTLGRKKAARKMAOTIRRLRGFLIRSPCE 785
Db 717 WKGLMQRKRYLAKRAOYIIMOSYCRKLAAQAAKKREADKITAFLKGFITRDARNGF 776
Qy 786 NAFLLIVKRSFLLNLRLPRLVLTSTPTPALREASLRELCKMNMWYKCSIS 845
Db 777 NEEFIANAKRMWLLRLAKELPTKVLKDSWPHAPCHCEASGLIHLRHILHARILYRLKT 836
Qy 846 PEMWQOLOOKAVASEIFKGGKNDYPOSVRLFISTRIGTEI---SPRVLOSLOSEPQY 902
Db 837 POORKELEKVLAKVFKGKKNNYASSVTFQEDRLPKHEIOWNDPFASTFESQELKY 896
Qy 903 AVPVYKDRKGKPRRQQLLTTPSAVYIVEDAKVKQ-----RIDYANLTGISVSSLD 955
Db 897 QSFCTKFDHRGYSRDEPILLSNKAIYVLDGKTYKOKHRLPLDKIDF-----TLTHND 950
Qy 956 SLVPLHQREDNOKGQVYVLOSQSHVETLTKTALSADRVNNINI--NOCSTFAGCPGRGD 1014
Db 951 DLAWIRLPLDKDKGDLILIRLIEFSTYIIDVTGASIVSVDRNSLEHNWVKGKG 1010
Qy 1015 IIDTSGSELLTKAKNGHLAV 1037
Db 1011 VIDOTGAEPGVVRDK-CHLVII 1032

```

```

RESULT 7
ID 063357 PRELIMINARY: PRT: 1006 AA.
AC 063357
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Myosin I.
GN MYR4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON NCBI_TaxID=10116;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=94308268; PubMed=8034741;
RA Boehler M., Kroschewski R., Stoeckli H.E., Behrmann T.;
RT "Rat myr4 defines a novel subclass of myosin I: Identification,
RT distribution, localization, and mapping of calmodulin-binding sites
RT with differential calcium sensitivity.";
RL J. Cell Biol. 126:375-389(1994).
DR EMBL: X71997; CAA50871.1;
DR HSSP: P08799; 1MND.
DR InterPro: IPR000048; IQ_region.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00612; IQ; 2.
DR Pfam: PF00063; myosin_head; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR PRODOM: PD000355; myosin_head; 1.
DR SMART: SM00015; IQ; 1.
DR SMART: SM00242; MYSC; 1.
SQ SEQUENCE 1006 AA; 116137 MW; 9AAB626A0FDA2A CRC64;

```

```

Query Match 34.5%; Score 1865; DB 11; Length 1006;
Beet Local Similarity 41.7%; Pred. No. 8,5e-120;
Matches 426; Conservative 158; Mismatches 301; Incls 136; Gaps 26;

```

```

Qy 28 GVQDVFLLNFTSEAFINLRRRPRENLITYIGVLTAVSDYRDLQIYSRQHMERVYG 87
Db 10 GAADEVLTMDT-VSMPEFMANLRLREFKGIYIFIGEVVSVNPKVNLITGKDTIEQYKG 68

```

```

Qy 88 VSPFVPPHFLFAVDIYRRLTERRDQAVMIGESGAGCTEATKRLLOFYAETCAPER 147
Db 69 RELTERPPHFLFALIDAAYKAMKRRSKDTCIMISGSGAGTEASKYIMQYIAITPNSQ 128
Qy 148 G--GAVDRLLQSNPVLEAFGNKTLNDNSSPFGKMDYQDFKAPVCGHLLSVL 205
Db 129 AELERVANMLLSKCVLEAFGNKTNNDNSRFGKMTDNEPFGKDPGGHINNVLK 188
Qy 206 SRVYHONHGERNPFVYQLLEGGEETLRILGLERNPOSYLYLVKQCAVSSINDKSD 265
Db 189 SRVYQGPGRSFHSFVQLQGGSEQLHSLHLOKSLSSVNYIRVAGOLK-SSTINAAE 247
Qy 266 KVMKALSYIDFTEDVEDLLSYASVTLHNGHFAADES---NAQVTEQMLYLR 321
Db 248 KVAADAKVIGFKEPVEIQYKILAILHGNKFLYDGDGTPLIEGKV----VSVAE 302
Qy 322 LIGVEGTLREALTHRIKAGEELSPNLQ--AAVARDALKAAYSRTFWLVKIR 380
Db 303 LLSFKADMVERKALLYR-TVATGRDIIIDKQHTBQASYGRAFAKALYERLFCYITRIND 361
Qy 381 SLASKDAESPMSRSTVYLLDLYGFEVQHNSEFQFCINYNCEKLDQFTELTLKSEQ 440
Db 362 IIEVKNYDTVHGKNVTYIGVLDIYGFIEDNNSFEQFCINYNCEKLDQFTELTLKSEQ 421
Qy 441 EYEAEIAMEPVOYFNKKIICDLVEEKFGIISIIDEECLRGAEATDLFLEKLEDTVP 500
Db 422 EYREGIPMKHIDYFNQIIVDLVEQOHGIIALIDDACMNVGKVTGMEFLALNSL 481
Qy 501 HPHFLTKLADOKTRKSLDGEFRLHYAGEVYVTVGFLDKNDLIFNLKETMCSMN 560
Db 482 HGFESSRKTCASDKILLEPR-DFRIHAGDVYVYIGFLDKNKDLPDFEFLKMYNSN 540
Qy 561 PIAQCFDKSELS---DKRPEVATQFMSLQVLEILRSKPAYIRCIKPNDAKOPGR 617
Db 541 PVLKNMPEGLKSLITEVTRKPLAATLFFKNSMALVDNLASKPEYVVICIKPNDKSPQI 600
Qy 618 FDEVLRLHGVKYLGMENLRVRAGFAVRKYEAFLQRYKSLCEPFWMAARPODGVAV 677
Db 601 FDDERCRHOVEYIGLEENVRVRAGFAFRQYIEKFLRYKMSSELPWPN-HDLPDKEAV 659
Qy 678 --LVRLHGYKPEEKMGRTKIFIRPKTLFATEDSLEVRROSLATKIOAMRGFIWROKF 735
Db 660 KKLIERCGFO-DDVAVGKTKIFIRPKTLFLE---ELAQMLV----- 699.
Qy 736 LKVRKSAICISWMRGTLGRKKAARKMAOTIRRLRGFLIRHSPCEPENAFLDHYRA 795
Db 700 ---RVVLELQKVVNRGTLARMR-YKRTKAALTIIRYRYRYKV----- 737
Qy 796 SFLLNLRLQRL--PRNVLD-----TSWTPPALREASELRELCKMNMWYKCSISPEWK 849
Db 738 SYIHVARFRFHGVKNRDRGKHVKNTPPKVYLRREELAQISFNWKRASQLIKTIPASDL 797
Qy 850 QOLOOKAVASEIFKGGKNDYPOSVRLFISTRIGTEISPRVLOSLOSEPQYAVPV-- 907
Db 798 PQVRAKVAAMEMLKGORADL--GLOAMEGNYLASKPDP--QTSGT-----FVYVANE 847
Qy 908 -----KY-----DKGKTPPRQLLTTPSAVYIVEDAKV 936
Db 848 LKRKQKYNVLFSCVRYKVNRESKVEDRAIFVTDRHLKMP-----TKQYKV 895
Qy 937 KORIDYANLTGISVSSLSFLVLRVQREDNOKGQVYVLOSQSHVETLTKTALSADRVNN 996
Db 896 MKTILPLYNLTGISVNGKQDLVVFHT--KDNK-----DLIVCLFSKOPTHESRIGE 944
Qy 997 I 997
Db 945 L 945

```

```

RESULT 8
Q27328 PRELIMINARY: PRT: 1017 AA.
ID Q27328

```

AC Q27328.  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE HUM-5 protein.  
GN MCR1A.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Lloyd C.;  
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
RP SEQUENCE FROM N.A.  
RA Cope M.J.T.V., Kendrick-Jones J.;  
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z35603; CAAB4673.1; -;  
DR EMBL; X75564; CA53244.1; -;  
DR HSSP; P08799; 1MND.  
DR InterPro; IPR000048; IO\_region.  
DR InterPro; IPR001609; myosin\_head.  
DR Pfam; PF00612; IQ; 1.  
DR Pfam; PF00063; myosin\_head; 1.  
DR PRINTS; PR00193; MYOSINHEAVY.  
DR ProDom; PD000355; myosin\_head; 1.  
DR SMART; SM00015; IQ; 1.  
DR SMART; SM00242; MYSC; 1.  
KM myosin.  
SQ SEQUENCE 1017 AA; 116557 MW; 461FF63A2C955ED8 CRC64;  
  
Query Match 32.5%; Score 1756; DB 5; Length 1017;  
Best Local Similarity 39.1%; Pred. No. 2,9e-112;  
Matches 405; Conservative 179; Mismatches 329; Indels 134; Gaps 24;  
  
QY 28 GVQDFVLLNFTSEAFTEENLRFRRENLIYTYIGPVLSVNPYRDLQIYSHOMEERYRG 87  
DB 12 GVEDVLTLSTDLKSV-VQNLQLRFOKGRITYYIGEVLAANPYRQLGIYEKSTVDYQKG 70  
  
QY 88 VSFYEVPLFLAVNATYVRALRTERDDAVMSGSGAGKPTATARKLLOFYAETCPAPER 147  
DB 71 REIYRAHPVAIAADAAARSMKRFGRDSCIYSGSGAGKETSRTIKMYLAATINVRQ 130  
  
QY 148 G--GAVPRDLLOSNPVLEAFNAKTLRNDNSRFQGVDFGAPGHIISYLEEK 205  
DB 131 GEIEVKKVLNLSKNCILAFGCAKTRNDNSRFCKYMHINFDYGDYGVGINSYLLLEK 190  
  
QY 206 SRVYQHNGERNFHYQLLEGGEETLRILGLEERNPQSYLYLVKQCAKYSINDKSDW 265  
DB 191 SRVYRQGEERNFHYQLVNGDGLLRQFGLTRKADQYVLYLNGQSHKVASINDSRDF 250  
  
QY 266 KVMKRAL-SVIDFTDEVEDLSTIYASVLHGINHPAEDDSN--AQYTTEQLYLRRL 322  
DB 251 AEVQALASIHFTFDQVDESMVSIAGLIHLGNVRFIDGENSSGAVHIAEKALONARC 310  
  
QY 323 LGVEGTTLRREALTHRKITLAKGELLSPNLBEQAAVAPALAAVSRPTMLVRKINSL 382  
DB 311 LNVITDELAKSLSSQVVAHGDIVAKQHDVANAAYTRDLAKALERLEFSMWVSVYNEAT 370  
  
QY 383 ASKDAESPMSRSTVVLGLDLYIGFEVFOHNSFEOPCINVCNEKLOOLFETLTKSEOEY 442  
DB 371 SVQN--SSRYSKSHVIGVLDYIGFEIFGNSPEOLCINVCNEKLOOLFETLTKSEOEY 428  
  
QY 443 EAEGTAMPEVOYFNKKITCDLVEEKFKGISLDEECLEPGATDLTLEKIEDTVKPRHP 502  
DB 429 EREGTKVWKIEFNNKVICDLVEIPRTGILSLDEACASIGAVTQKVFGLGELDKLKSHK 488  
  
QY 503 HFLTKLADQTKRSLDGEFRLHLYAGFVTVSGTGLDKNDLDFRNKTKETMSMNP 562  
DB 489 HTSRLN--KQSDKSMGEFEFKITHYAGDVTYSVNGFMDKNDTLFQDLKRLIYHSKRL 546  
  
QY 563 MAQCF--DKSELSDKKRPETVATQFKMSLLQLVLEILRSKEPAYTRCIKPNDAKQGRPD 619

DB 547 VKSLPPDSKSAEVAENRRPPTAGFLFKNSMSELVQLAQKEPHYIRCIKPNEEKSNVFD 606  
QY 620 ELLIRHOYKYLGMENLVRVRAGFAVRKYEAFILORYSLCEIYPMNAGR-----OD 673  
DB 607 LERVHORYLGLLENVRNRAGFAHRMPYDRFVNRKILCASTWP---NPRQOQLD 662  
QY 674 GVAVAVRHLYGKPEEYKMGORTKIFRFPKTLFPATEDSLVRROSLATKIOAAMRGFHWQ 733  
DB 663 SCMQLLESAGL-AQDCVQGRTKIFIRSPQTVRELELRTEQLPNVITTEQKKVVRGQORE 721  
QY 734 KFLRVKRSATCIQSWRRGTGLGRRAKRRKMAOTIRLRIGTILRHSPRCPENAEFLDHV 793  
DB 722 RY-----RRMLAVRKITGAYRRYKLSYIWO----- 747  
QY 794 RASFLNLRROLPRNVLDTS-----WTPPEPALREASLELRELCMNMYKCRSIS---P 846  
DB 748 ----LINAERDY-RKMRDICKSIRWPAPELVLAQVRSIRVRWHQ---WRAATILARMP 799  
QY 847 EWKQOLOOKAVASELFGKCKDN-----YPOSVPLFISTRLGTEISPRVLOSIG 896  
DB 800 HLRASLPQKIAFELVNKKKEWGYTRMRG DYLSQGELELPTVSTYHDIQALRQ-- 857  
QY 897 SEP---IQYAVPVKYYDRKGYKPRPQLILPSSAVYVEDA---VKORIDYANLTGISV 950  
DB 858 SHPFCKVLFSTYVQKFN-KFNKSSLRVILVTRFVAKLENKKFKLKEPIPLQISIRISV 916  
QY 951 SLSLSLFLVLYHQRD-----KKQGDVYLOS DHVTE 982  
DB 917 CAESNGLEFVHYGDNNDIVGCAKNTNEERVGEMIGTLAHHYDKITMRSPVLIOG-AVVC 975  
QY 983 TL---TKTALSADRVNN 996  
DB 976 TLGGKTRITRVDAENN 992  
  
RESULT 9  
Q90748 PRELIMINARY: PRT: 1099 AA.  
ID 090748  
DC 090748:  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Brush border myosin IB.  
GN CBRMB.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauaria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Knight A.E., Kendrick-Jones J.;  
RT A novel vertebrate myosin I.\*;  
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
DR EMBL; X70400; CAA49850.1; -;  
DR HSSP; P08799; 1MND.  
DR InterPro; IPR000048; IO\_region.  
DR InterPro; IPR001609; myosin\_head.  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF00612; IQ; 1.  
DR Pfam; PF00063; myosin\_head; 1.  
DR Pfam; PF00018; SH3; 1.  
DR PRINTS; PR00193; MYOSINHEAVY.  
DR ProDom; PD000066; SH3; 1.  
DR ProDom; PD000355; myosin\_head; 1.  
DR SMART; SM00015; IQ; 1.  
DR SMART; SM00242; MYSC; 1.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS00002; SH3; 1.

<b>RESULT 10</b>					
ID	012965	PRELIMINARY;	PRT;	1109 AA.	
AC	Q12965;				
DT	01-NOV-1996	(TREMBLrel. 01, Created)			
DT	01-NOV-1996	(TREMBLrel. 01, last sequence update)			
DT	01-JUN-2002	(TREMBLrel. 21, last annotation update)			
DE	Myosin-IC.				
OS	Homo sapiens (human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE OF 1-245 FROM N.A.				
RX	MEDLINE=95018277; PubMed=7932763;				
RA	Bement W.M.; Witt J.A.; Mooseker M.S.;				
RT	"Cloning and mRNA expression of human unconventional myosin-IC. A				
RT	homologue of amoeboid myosins-I with a single IQ motif and an SH3				
domain".;					
RL	J. Mol. Biol. 243:356-363(1994).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	Bement W.M.;				
RL	Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.				
CC	-1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.				
EMBL:	U14391; AAA62667.1; "				
HSSP:	P08799; IMD.				
InterPro:	IPR000048; IQ_region.				
InterPro:	IPR001609; myosin_head.				
InterPro:	IPR001452; SH3.				
Pfam:	PF00612; IQ_1.				
Pfam:	PF00063; myosin_head; 1.				
Pfam:	PF00018; SH3; 1.				
PRINTS:	PRO0193; MYOSINHAIV.				
PRINTS:	PR00452; SH3DOMAIN.				
ProbDom:	PDO00066; SH3; 1.				
ProbDom:	PDO00355; myosin_head; 1.				
SMART:	SM00242; MYSC; 1.				
DR	SMART; SM00326; SH3; 1.				
DR	PROSITE; P55002; SH3; 1.				
KW	SH3 domain.				
SQ	SEQUENCE 1109 AA; 127040 MW; 0C1C74BAA3705905 CRC64;				
Query Match		31.8%; Score 1715; DB 4; Length 1109;			
Best Local Similarity		37.4%; Pred. No. 2.2e-109;			
Matches 398; Conservative 172; Mismatches 295; Indels 198; Gaps					
24,					
OY	28	GVDGFLELNFNFSAEAFINLRKRREPNLIYYIGPVLSVNPYRDQLIYSROHMERARG	87		
		: : :     : : : : :             : : : : :     :			
Db	20	GVDADMWLKSTI-EENSVLENKKRYMDYITFYIGSLVSVPFKNPYGGEKEIEMTQG	78		
OY	88	VSEFEVPHLEAVADVTYRALRTERDDQAMTSGESGAGCTEATRLDFOYAETCPAPER	147		
		: :         :			
Db	79	AAQVENPHPIVALANMNYMIIDENCVIITSGESGAGKTVAARYIMSYISRVGGGKV	138		
OY	148	GGAVNDRLLQSPPVLEAFENCAKTLINDNSRRCKGYMDQFDPKGAPVGCHIIISYLESKR	207		
		: :         :			
Db	139	VGVAVDDIIQSPPLEAFEGMAKTVNNNSSRFKGYFEIQSPGGEPSDGKISNFLEKS R	198		
OY	208	VHQHGNERNFVFQOLLGEGEEETFLRLRGLEERNQSYLVLYKGCACAKYSSINDSKMV	267		
		:                         : : : : :     : : :     : : :     :			
Db	199	VVMRNPGERSFFIIFQLLEGASAECHKSLGI-TSMNDYIYLISLSSSYKVVDDIDRKREFQE	257		
OY	268	MKRALSYIDFTDEVEDDLSTIYASVLHLGNTHFAADEDSNAOVTTENOLKYITRLLGVEG	327		
		:     : : : : :			
Db	258	TJHANVNIYIAEEDTLVLIQTIGILIGNISF-KEVGNVAAVEEELPAFPAYLIGVQ	316		



Qy	677	VLVRLHGYRPEEKMGKRTGIFIRFKPLTEPATEDSLVEVRQSLATKIQAMRGHNRKQKTL	736
Db	655	HLIRAVNMEPDQIQMSKISTVAVFNKPSLFLLEVEERKFPDGFARTIQKAW	704
Qy	737	RKRSACIQSWMRGTLGRKRAKAKRRMAQGITRILRIRGLRHSRPPENAFLLDIYRAS	796
Db	705	-----RRHVAVRKY	713
Qy	797	FLNLRLRQLPRVNLQTSWTFPPRLRREASSELLRELCKMKNMYKCRSISPEMKQQLQKA	856
Db	714	-----EEMRE	718
Qy	857	VASELPKGRKNDYPOSVRLFTSTRLTGTEELSPRYLSLG-SPEIYAVPVVYDKRGYK	915
Db	719	EASNILLNKERRRNSINKNFVGDLGLEF-RPELRQFLGKKEKVDYDSDVYTKYDR-RK	776
Qy	916	PRRQOLLTPSNAVIVEDAK-----VKQRIDYANILGTISVSSLSLSLFLVHV	962
Db	777	PIKRIOLLTPKCVVYIGREKKMKKGPEQKQVCEVKKKKYDIALRGVSLSTRQDDFL--	834
Qy	963	QREDNKKQKGDVYLOSQDHYETLTTRK-----ALSADVNNININQSGSYFA	1007
Db	835	-QED-----ADSFLEYSVETEPYSLCKKFEATRRPLPLTFSDTLQFVRKKEG	884
Qy	1008	GEPGRDGIIDFTSGSELLITAKNGHLAV	1036
Db	885	GGGGRSVTFSGFGDLAVLKVGGRTLTIV	913
RESULT	12		
063356			
ID	063356	PRELIMINARY:	PRT: 1107 AA.
AC	063356:		
DT	01-NOV-1996	(TREMBLrel, 01, Created)	
DT	01-NOV-1996	(TREMBLrel, 01, Last sequence update)	
DT	01-JUN-2002	(TREMBLrel, 21, Last annotation update)	
DE	Myosin I heavy chain.		
GN	MYR3.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_Taxid:10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-SPRAGUE-DAWLEY.		
RX	MEDLINE-95247829; PubMed-7730414;		
RA	Stoeckli H.E., Ruppert C., Reinhard J., Bahler M.;		
RT	"A novel mammalian myosin I from rat with an SH3 domain localizes to		
KT	Con A-inducible, F-actin-rich structures at cell-cell contacts.";		
RL	J. Cell Biol. 129:819-830(1995).		
CC	-1-SIMILARITY: CONTAINS 1 SH3 DOMAIN.		
CC	EMBL: X74815; CAA52815.1; -		
DR	HSSP: P08799; 1IND.		
DR	InterPro: IPR000048; IQ_region.		
DR	InterPro: IPR001609; myosin_head.		
DR	InterPro: IPR001452; SH3.		
DR	Pfam: PF00612; IQ.1.		
DR	Pfam: PF00063; myosin_head.1.		
DR	Pfam: PF00018; SH3.1		
DR	PRINTS: PR00193; MYOSINHEAVY.		
DR	PRINTS: PR00452; SH3DOMAIN.		
DR	ProDom: PD000066; SH3.1.		
DR	ProDom: PD000355; myosin_head.1.		
DR	SMART: SM00242; MYSC.1.		
DR	SMART: SM00326; SH3.1.		
DR	PROSITE: PS50002; SH3.1.		
DR	SH3 domain.		
QO	SEQUENCE 1107 AA: 12682 MW: B9DBFB80CE047148 CRC64:		

QY	28	GVOEVLLENTSPAATIELRRREPEENILYNYIGVLYSVNPRYDLOIYSKOMEHTRG	87
Db	20	GVDDMVLKSKIT--ESSIVENLKKRYMDVITFYIGVLYSVNPFKOMPYPEKEKETEIMOG	78
QY	88	VSFEYVPHLEAVADVYVVALRTERRDOAVMISGESGAGKETATKRLLOFAETCPAPER	147
Db	79	AAOYENPHIYALADSKYRNMIDRENOCVILISGESGAGKTVAAKYIMSYSVSGSGPK	138
QY	148	GGAVRDOLLOSNVLEAFNAKTLRNDNSRRGKTYMDQFDEKAPVCGHILSTLEKSR	207
Db	139	VQHKDILLOSNVLEAFNAKTVRNNNSSRGKFEYQFSGPEBDGKISNLEKSR	198
QY	208	VHQHNGERNHFVYOLLEBGEBEETLRJRLGERNQSYLYLVKGCAVSSINKSMPK	267
Db	199	VVMNRGESEFHIFYOLIGASPEQOSLGI--TSMDDYYILSTLSGSXYVDIDDKRPOE	257
QY	268	MKRALSYIDTBEVEDLSTIYASVYLHGNIHFADEDSMAQVTTENOLKYTRLLGVEG	327
Db	258	TLHAMNYIGFSEEQPLVLOIYAGLHGNINF--KEVONYAAVSEBEFLAFPAVLLOQ	316
QY	328	TYREALTHRKILA---KGELLSPNLDEOAAVYARDAKAAVSRFTMYLVRKINSLA	383
Db	317	DRLEKRLTQNSDKSKGSGSESINHTLVNQACYTRDIALAALHARVFDLVDSINKAM--	375
QY	384	SKRAEBSMRSTVLCLODIYFEYVFOHNSFEQFCINCYNEKLOQLEFELTSLSEDEYE	443
Db	376	EKHEEYN-----IGVLDYGFELFOKNNGFEQFCINFEVNEKLOQIFELTSLKAEQEYV	429
QY	444	AEGIAEPVOYFNKKIICDLYVEKERK--GIISILDEBL---RGEATDULFLEKEDY	498
Db	430	OEGIRMTPEYENNNKIYCDLIESKNPFGIMSILDDVCATGMHAYEGADOTLOKLOMOI	489
QY	499	KHPHPLTKLADOKTRKSLDGBERLLHAYGEVYTVSYGLDKNNDLRLNKETMCS	558
Db	490	GSHEHF-----NSMNG--FIHHYAGKYSYMDGCEBERNDVLFMDLIELMOSS	537
QY	559	MNEIMQCFDKSELSDRK--REPVTATQKMSLLOLVLELSKBEAVYRCKIKPDAQOPGR	617
Db	538	ELRFLKSLPEHLQADKKGRPTTAGSKTKKQANDVSTLKKCTPHYIRCIKPEYTKKPAD	597
QY	618	FDEVLIRHOVKYLGIMENLYVRAGFAVRRKYEAFLORYASLCPETPMAGRPDQVAV	677
Db	598	WESERKIQOVELYGLKENIYVRAGVAYRVEYKFORVAILTKAVPMVRQDEKGVHL	657
QY	678	LYRHLGYRPEEKMKRTEKIFIRPKPTLFAPTMDSLAVKROSLARKIOAMRGHMKFLR	737
Db	658	LLOSVNMSDOQOLRSKVFYFKABESJFLLEEMKERYDYGAIVOKYWRKVFARKKYQ	717
QY	738	VKRSACIQISMMWGTGLGRKAARKMAAQITRLIRGFIHLHSPRCPENAFFLDHVRSF	797
Db	718	REED-----	721
QY	798	LLNLRLQLEPRNVLDTSMTPPPALREASSELLRELCMKNMWYKYSISEMKOOLQORAV	857
Db	722	-----	721
QY	858	ASEIFKGGKNDVPOSVPLRFTSTRLTGEESIPRYLOSLS--BPLOYAVPYVYDGRGYK	916
Db	722	ASDLLNKKEKRRNSINRNFIQDYGIMHE--HPELOQFVGKREKIDPADIYTYDNR--FRG	779
QY	917	RPROLLTSPSAVJIEADKAKQ-----RIDVANTLGISVSSLSDLSLFLVHQ	963
Db	780	VKRDILPKCLYLTIGREKQYOGDKGLYEVLRKRIEVERILSYASTLMQODIFLMO	839
QY	964	REDNKOKGDVULQSDHYIEFTLTK---ALSADVNNMINOGSITFAG-----G	1009
Db	840	-----EIDSILSEYKTEPISLTLTKRYEKTOKOLPLKFSNTLELKLRKNMG	887
QY	1010	P---GRDGIIDFTSG--SELLITRAKNHILAV--VAPRL--NSR	1044
Db	888	PWSAGSGROYOFYGGFGLDILTKPSKNVLOVYSIGPGLPKNAR	929

RESULT 13

09NGL7 PRELIMINARY; PRT: 1053 AA.

ID 09NGL7

AC 09NGL7

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE Amoeboid myosin I (Fragment).

OS Strongylocentrotus purpuratus (Purple sea urchin).

CC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoida; Echinoidea; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus.

CC Strongylocentrotus.

CC NCB1;taxid=7668;

OX NCB1

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20440049; PubMed=10984411;

RA Sitotkin V., Seipel S., Krendel M., Bonder E.M.;

RT Characterization of sea urchin unconventional myosins and analysis of

RT their patterns of expression during early embryogenesis.\*;

RL Mol. Reprod. Dev. 57:111-126(2000).

DR EMBL: AF248488; AAF71717.1; .

DR HSSP: P08799; 1MD.

DR InterPro: IPR001609; myosin\_head.

DR Pfam: PF00063; myosin\_head.1.

DR PRINTS: PR00193; MYOSINHEAVY.

DR ProDom: PD000355; myosin\_head.1.

DR SMART: SM00242; MYSC.1.

FT NON\_TER 1053 1053

FT SEQUENCE 1053 AA; 119723 MW; 0B31468B9F9FD03C CRC64;

Query Match 31.4%; Score 1695; DB 5; Length 1053;

Best Local Similarity 37.4%; Pred. No. 4.9e-108;

Matches 400; Conservative 171; Mismatches 291; Indels 208; Gaps 24;

QY 24 RDRVGOEVLLENTSEAPFLENRRPRENLITTYIGPVLVSNPRDLOIYSRQHE 83

DB 10 RRTSGVEEDVLLSKI-OGCAIVENLKRFRMDLIYTYIGPVLSINPKOMPEFTEKEVD 68

QY 84 RFRGSFVEPHLLFAVVDVYRALTERRDOAVMISGSGSGKTEATRKLLQFYAETCP 143

DB 69 MYGAAITENPHYVALTDDMRNMLIEYNOCVITISGSGAGTAAKYIYATYAKV-- 126

QY 144 APERGCA---VRDLLOSPLYLEAFNAKTLRNDSSRFKYMVOEFPFGAPVGHIL 199

DB 127 --SGGATVQHKVEYIQSINPLLEAFNAKYRNNSSEFGKYVLEIQTRACAPRGKIT 184

QY 200 SYLLEKSRVYHONGERNFHYVQLLEGGEETLRLRLGNRPQSYLYLVKGCCKAVSSI 259

DB 185 NFLLEKSRVYHONGERNFHYVQLLEGGEETLRLRLGNRPQSYLYLVKGCCKAVSSI 243

QY 260 NDKSMKVMKRALSYIDPFDEVEDLISVSLHNGHFAADSDNSAQVTENQKYL 319

DB 244 DVVKEYKDTLNAASYIGISEDQENLVSCVGLIHLGNLVF--VERDNTAVIHIDDEFLDP 302

QY 330 TRLLGVEGTLREALTHKRIIA---KGEELSPNLDEQAAYARDALAKAVYSRTFTWLV 375

DB 303 SYLLGLDEALREALKKILSRMDSKMGKAEHIEVLTINTEQAATRALAKAVASRLFDLV 362

QY 376 RKINSLASKDAESPMSSTVYGLDITYGFVFOHNSFEOPCINCNKLOOLFIELTL 435

DB 363 SSINAMOKOKE-----ITIGVLDIYGFELFOONGEOPCINFNVEKLOOLFIELTL 415

QY 436 KSEGEYEAGEJAMPEVYFNKKIICDLVEEKF-KGIISIDEECL---RPEADITFL 491

DB 416 KAEQEEYIAEGTKMKKIDYFNKKIYCDLLESKVPVPGVMLDDVCATMNAASDGADEKIL 475

QY 492 EKLEDTVRPHPLTHKLADQKTRKSLDGEFRLLHYAGEVTVSYVGFGLDKNNDELFRNL 551

DB 476 QKMNSAVGTHQY-----GVLSGFIYHHYAGQVYTVYEGFERRDVLFIDL 523

QY 552 KETMSSNMPIMAGCF-DKSELSDKKRPETVATQFRMSLDLVEILRSKEPAYIRCIKPN 610

DB 524 IELMOSSSENSPFMTLFPFKIAGQAKSRPTTASSTKIKTOANKLVKLMQSTPHYIRCIKPN 583

QY 611 DAKQPGREDEVLIRHQVYVLGIMENLRYRAGFAVRRKYEAFLQYKSLCPETPMWAGR 670

DB 584 ETKRPQDMEEKRVYHQVEVLGKLENIRVRAGFAVRRYEDKRLRRYALITPPTWRMGSD 643

QY 671 PODGVAIVLRHLGYKPEEYKMGRTKIFTRPKTLFATEDSLSEVRQSLATKIOAAWGFH 730

DB 644 VSGVNHLLRAVNMOTDFEQLGKTKVFTRKPSFLFLEQREKRFDFHARIIQKAFQFN 703

QY 731 WRQKFLYKRSALICQSMWRGTLGRRKAKRWAKQITRRLRGLRSLRSPRCENAPFL 790

DB 704 AOKYTLRLKQO----- 714

QY 791 DHVRSFLLNLRQLPRNVLDTSMPPTPALREASELLRELCKMNMVKKYCRISPENKQ 850

DB 715 ----- 714

QY 851 QLOQAVASEIFKGRKNDYPOSVRLFTSTRLGT-ELSPVQLSGS--EPIQYAVPV 907

DB 715 -----ASDILFDKERRRYSLSNENFGDYIGFPHNVS---LQALVARREIRFADKVN 764

QY 908 KYDRKGYKPRPQLLLTFSAAVYV--EDAK-----VKORIDVANLTGISVSSLS 954

DB 765 KYDR-KFVTKRDLLSGRNYTLIGREQIKGPEKGQIRDVYKRRIRPLNHIHVSYSLTQ 823

QY 955 DSLFVLHYQREDNKQKGVNLOSDHVIETLT---KTALSADRVNNININQG----- 1002

DB 824 DMFVLHYSEE-----YDSILESTFKTEFLYLNKKKFFQETLQKLMVNNDSIEFSYKKE 878

QY 1003 -----STFAGPGRGDIIDFTSGSELLITKAKNGHLAVAPRLNSR 1044

DB 879 GWGCGGFTVRFAQCGDPTAIL--KPSSKILTVSIGS-----LBR-NSR 920

RESULT 14

061080 PRELIMINARY; PRT: 1186 AA.

ID 061080

AC 061080

DT 01-AUG-1998 (TREMBLrel. 07, Created)

DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Myosin IC heavy chain.

GN MICH.

OS Acanthamoeba castellanii (Amoeba).

CC Eukaryota; Acanthamoebidae; Acanthamoeba.

OX NCB1;taxid=5755;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=88016163; PubMed=3477803;

RA Jung G., Korn E.D., Hammer J.A. III;

RT "The heavy chain of Acanthamoeba myosin IB is a fusion of myosin-like and non-myosin-like sequences.\*;

RT Proc. Natl. Acad. Sci. U.S.A. 84:6720-6724(1987).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=99079990;

RA Wang Z.Y., Wang F., Sellers J.R., Korn E.D., Hammer J.A. III;

RT "Analysis of the regulatory phosphorylation site in Acanthamoeba myosin IC by using site-directed mutagenesis.\*;

RT Proc. Natl. Acad. Sci. U.S.A. 95:15200-15205(1998).

CC -I- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

EMBL: AF051353; AAC98089.1; .

DR HSSP: P08799; 1MD.

DR InterPro: IPR001609; myosin\_head.

DR InterPro: IPR001452; SH3.

DR Pfam: PF00063; myosin\_head.1.

DR Pfam: PF00018; SH3.1.

DR PRINTS: PR00193; MYOSINHEAVY.

DR PRINTS: PR00452; SH3DOMAIN.

DR ProDom: PD000066; SH3.1.

DR ProDom: PD000355; myosin\_head.1.

DR SMART: SM00242; MYSC.1.

231

QY	959	VLHQREDNKGQDVYLIQSHVITFLTKAL-----SADRVNNININGSI	1004
Db	834	LISVNGHS-----NLCECRKRTTELIGVLKHPNSVRIQFADTFN-----V	874
QY	1005	TFAGC-----PQDGIIFTSGS	1022
Db	875	TLKGKTCVVFIRDPQGGDKVAGTKVVS	903
RESULT 15			
ID	Q19901	PRELIMINARY	PRT; 1100 AA.
AC	Q19901		
DT	01-NOV-1996	(TEMBLrel. 01, Created)	
DT	01-NOV-1996	(TEMBLrel. 01, Last sequence update)	
DT	01-JUN-2002	(TEMBLrel. 21, Last annotation update)	
DE	P29D10.4	protein.	
GN	P29D10.4		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;		
OC	Rhabditidae; Peloderinae; Caenorhabditis.		
OX	NCBI_TaxId=6239;		
RA	RP	SEQUENCE FROM N.A.	
RA	white S.;		
RL	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.		
RN	(2)		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=9069613; PubMed=9851916;		
RA	none;		
RT	"Genome sequence of the nematode C.elegans: A platform for		
RT	investigating biology."		
RL	Science 282:2012-2018(1998).		
CC	-i- SIMILARITY: CONTAINS 1 SH3 DOMAIN.		
EMBL	275952; CAB00095.1; -.		
DR	HSSP: P08799; 1MND.		
DR	InterPro: IPR000048; IQ_region.		
DR	InterPro: IPR001609; myosin_head.		
DR	InterPro: IPR001452; SH3.		
DR	InterPro: IPR000169; SHprot_acsite.		
DR	Pfam: PF00612; IQ; 1.		
DR	Pfam: PF00063; myosin_head; 1.		
DR	Pfam: PF00018; SH3; 1.		
DR	PRINTS: PRO0193; MYOSINHEAVY.		
DR	ProDom: PD000066; SH3; 1.		
DR	ProDom: PD000355; myosin_head; 1.		
DR	SMART: SM00242; MYSC; 1.		
DR	SMART: SM00326; SH3; 1.		
DR	PROSITE: PS50002; SH3; 1.		
DR	PROSITE: PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.		
KW	SH3 domain.		
SQ	SEQUENCE 1100 AA; 124885 MW; 2CDB6F02EDEDED9 CRC64;		
Query Match 29.0%; Score 1564; DB 5; Length 1100;			
Best local Similarity 34.7%; Pred. No. 5.7e-99;			
Matches 367; Conservative 172; Mismatches 320; Indels 200; Gaps 21;			
QY	27	VGQVDFVLENTTSAAETLENRRFRRELYTYTGVPVLSVNPYRDQIYSRQNMERYR	86
Db	14	VGVDMDVLLPKLT-EQSIVENIKKQLQNSITFTYIGPVLISVNPFKMPYLFKEKMLLYQ	72
QY	87	GVSEYEVPHLFAVADIVYALRTERRDQAVMISGSGAGTEATKRILOFYAETCPAPE	146
Db	73	GAAQENAPHTIALADNMVYRMMLIDNESQCYLISGSGAGCTVNAKFLMNTVLSRSGGQ	132
QY	147	RGCAVRRDLLQSNPYLEFNGAKTLIRNDSSRFQKYMVDYDFGAPVGHILSTLEKS	206
Db	133	KVQHKIDVLIQSNPLLEFNGSATVVRMNSSRFQKYEIVFSRSGEPVIGKTLISNPLEKS	192
QY	207	RVHQNHQERNEHFVYQLLEGSEETLRLGLERNPOSYFLVAGQCAKVSINDKSPMK	266
Db	193	RVHQNHQERNEHFVYQIACADKNIIRSTETI-GTGYVAVTINSCVYKATVATMDKQVET	266

```
QY 267 VAKKALSVIDFEDEVEDLSTIVASVLHGNHFAADEDSNAQVTTENOLKYLTRLLGVE 326
   1: 1: : : : 1: 1: 1: 1: 1: 1: : : : 1: : : 1: 1:
Db 252 STLHAKVYGVNDODOLEVLRIVATVHIGNITF--TEENNFAVSGKDYLEYPAFLGL 310
   : : : : : : : : : : : : : : : : : : : : : : : :
QY 327 GTTLREALTHRKIIAK----GELLSPNLNOAAYARDALAKAVSRTFTWLVRKINRSL 382
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 311 SADIEAKLTGRKMKESWGTOKEEDIMKLVNEQASYTRDAWKAIVARLFDYLVKKYNDAM 370
   : : : : : : : : : : : : : : : : : : : : : : : :
QY 383 ASKDAESPMSRSTTVLGLDIYGFVEVFOHNSFPOFCINVCNEKLOOLETELTLKSEQEEX 442
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 371 ---NITSQSTSDMSFSGIIDIYGFELFNNNGFPOFCINFVNEKLOQIFIELTLKAOEEX 427
   : : : : : : : : : : : : : : : : : : : : : : : :
QY 443 EAGCIAMEPVQYFNKKIICDLVEEK--KGIIISLDEBCLR---PGEATDLTLEKLEDIV 498
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 428 VREGIMTELDYFDNKKIVCDLITETKRPPGIMSLDPTCAQNHQREGVDPQLTTLTKSP 487
   : : : : : : : : : : : : : : : : : : : : : : : :
QY 499 KPHPHFTLTKLAQOKTRKSLDRGEFRLHAGVTVSVTGFLDKNNDDLFRNLKETMCS 558
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 488 AGHPHF-----GPGSDSEVYIKHYAGVTVTVNDGFCDRNRDVLTPDLLLMOKS 535
   : : : : : : : : : : : : : : : : : : : : : : : :
QY 559 MNPIMAOCPDKS--ELSDKKRPETVATOFKMSLQVLEILRSKEPAYIRCIKPNDAKOPGR 617
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 536 SRPFIALPPEENVAASGAKRPTFTSTKIRQANTIVESLMKCSPHYRCIKPNETKRPN 595
   : : : : : : : : : : : : : : : : : : : : : : : :
QY 618 FDEVLIIRHOVKYVGLMBNLVRRAGFAVRRKYAPLQRYKSLCPETPMMAWRPPOGVAY 677
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 596 WEESRYVKHQVEYLGRINIVRRAGFAVRRAPFAQRYAIVSPQTMPCFOGQORACEL 655
   : : : : : : : : : : : : : : : : : : : : : : : :
QY 678 LVNHLGKPEEYKMGRTKIFIRPKTLFATEDLEVRQSLATKIQAAWGFHWROKFLR 737
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 656 ICDSVMEKNQYOMGRTKLFVKNPESLFLEETREKFDGARVIOKAWROFSAR----- 710
   : : : : : : : : : : : : : : : : : : : : : : : :
QY 738 VKRSAICIGSMWRGTLGRKRAKRNMAOTIRRLINGFILRHSPRCPENAFILDHYRAS 797
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 711 ----- 710
   : : : : : : : : : : : : : : : : : : : : : : : :
QY 798 LNLRLRQLPRNVLDTSWPPPPALREASELLRELCKMNMWYKCRSISPWKQOLOKAY 857
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 711 -----KOHIKOKEQ 719
   : : : : : : : : : : : : : : : : : : : : : : : :
QY 858 ASLFIKGGKKNYPOSVPRLEFISTRIGTEISPRVLOSL--GSEPIQYAVPVVYKDRKG 915
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 720 AADLMYGKKERRRYSLRNRFVGDYIGLEH--HPTLOSLVGKRQVRLEACTANKYDRK-FR 776
   : : : : : : : : : : : : : : : : : : : : : : : :
QY 916 PRPROLLTPPSAVVIEDAKV-----KORIDYANLTGISVSSLSDSLFLVHY 962
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 777 VTKLDLLITVNNHLTLGKEKVKNGPEKGKIVEYIKRQFLDPQIKSIGLSIPYODDFVILYL 836
   : : : : : : : : : : : : : : : : : : : : : : : :
QY 963 QREDNKOQGDVVLQSDHVIETLTK---TALS---ADRVNN-----ININGST 1004
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 837 GNDYDS-----LLETPEKTECFALSKAYKERTNGTLHLDFRSSHVVSYKKMKF 886
   : : : : : : : : : : : : : : : : : : : : : : : :
QY 1005 TFAGP-----GRDGIIDFTSGSELITRKAKNGHLAVVA 1038
   : : : : : : : : : : : : : : : : : : : : : : : :
Db 887 DFDGKRTVOFGNDG---TSSAEK--TLKPNCKVLNVS 919
   : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: December 9, 2002, 16:46:44  
Job time : 89 secs



